



Essays on the Economics of Risk and Financial Markets

Citation

Turley, Robert Staffan. 2013. Essays on the Economics of Risk and Financial Markets. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11095955>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

*Essays on the Economics of
Risk and Financial Markets*

A DISSERTATION PRESENTED
BY
ROBERT STAFFAN TURLEY
TO
THE DEPARTMENT OF ECONOMICS AND THE HARVARD BUSINESS SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
BUSINESS ECONOMICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MARCH 2013

© 2013 - ROBERT STAFFAN TURLEY
ALL RIGHTS RESERVED.

Essays on the Economics of Risk and Financial Markets

ABSTRACT

Prices in financial markets are primarily driven by the interaction of risk and time. The returns to financial assets over long time horizons are primarily driven by fundamental news regarding their promised cash flows. In contrast, short-run price variation is associated with a large degree of predictable, transient investor trading behavior unrelated to fundamental prospects.

The quantity of long-run risk directly affects economic well-being, and its magnitude has varied significantly over the past century. The theoretical model presented here shows some success in quantifying the impact of news about future risks on asset prices. In particular, some investing strategies that appear to offer anomalously large returns are associated with high exposures to future long-run risks. The historical returns to these portfolios are partly a result of investors' distaste for assets whose worth declines when uncertainty increases.

The financial sector is tasked with pricing these risks in a way that properly allocates investment resources. Over the past thirty years, this sector has grown much more rapidly than the economy as a whole. As a result, asset prices appear to be more informative. However, the new information relates to short-term uncertainty, not long-run risk. This type of high-frequency information is unlikely to affect real investment in a way that would benefit broader economic growth.

Contents

1	PRICE COMOVEMENT AND TIME HORIZON: FADS AND FUNDAMENTALS	1
1.1	Modeling and Measuring Comovement	8
1.2	Short-Run and Long-Run Comovement in US Equities	16
1.3	A Regression Methodology for Correlations	21
1.4	Explaining Empirical Correlations	25
1.5	Implications for Asset Prices and Investors	40
2	AN INTERTEMPORAL CAPM WITH STOCHASTIC VOLATILITY	55
2.1	An Intertemporal Model with Stochastic Volatility	65
2.2	Predicting Aggregate Stock Returns and Volatility	81
2.3	Pricing Cash-flow, Discount-Rate, and Volatility Betas	104
2.4	The Contribution of Stochastic Volatility	134
3	INFORMATIVE PRICES AND THE COST OF CAPITAL MARKETS	138
3.1	A Stylized Model of Capital Markets	147
3.2	Explaining the Empirical Growth in Capital Market Spending	176
3.3	Market Activity and Asset Prices in the Cross Section	189
3.4	Implications and Conclusions	203
	REFERENCES	209

Listing of figures

1.1	Correlations for Heinz, Phillip Morris and Harley Davidson	3
1.2	Comparing Empirical and Reshuffled Correlation Differences	24
1.3	Annualized Volatility and Correlations for Risk Factors, 1970-2009	41
1.4	Event Time Returns to X_t Components of the L/S Portfolio	47
1.5	Annual Calendar Time Returns to L/S Portfolio	49
2.1	Approximate Gamma-Omega Relationship	76
2.2	Consumption Growth Variance and Risk Aversion	79
2.3	Realized and Expected Variance, 1926-2011	87
2.4	Normalized News Series	96
2.5	Key Components of Long-Horizon Volatility	103
2.6	Pricing 25 Size and Value Portfolios, Early Period	122
2.7	Pricing 25 Size and Value Portfolios, Modern Period	125
2.8	Contour Plots Showing Goodness-of-Fit	127
2.9	Pricing Risk Sorted Portfolios, Early Period	130
2.10	Pricing Risk Sorted Portfolios, Modern Period	133
2.11	Pricing Risk Sorted Portfolios, Modern Period	135
3.1	Capital Market Spending and Compensation	141
3.2	Contrasting Banking and Credit vs. Capital Market Activities	143
3.3	Intuition behind model equilibrium	155
3.4	NYSE Commission Schedule, 1956	178

3.5	Transaction Cost Time Series	182
3.6	Predicting the cost of capital markets using the cost of transacting .	186
3.7	Rolling Regression Coefficient and Moving Average, 1965-2010 . .	197
3.8	Welfare in the case of inelastic investment supply (short-horizon) . .	207
3.9	Welfare in the case of elastic investment supply (long-horizon) . . .	207

List of Tables

1.1	Data Coverage for Correlation Estimates	18
1.2	Summary Statistics for Correlation Estimates	19
1.3	Summary Statistics for Primary Explanatory Variables	26
1.4	Summary Statistics for Dummy Variables	29
1.5	Cross-Sectional Regressions of Correlation Difference, $\Delta\hat{\rho}_{ij}$	31
1.6	Microstructure Robust Correlation Differences	35
1.7	Regressions Adjusted for Microstructure Effects, $\Delta\dot{\rho}_{ij}$	36
1.8	Regressions of $\Delta\dot{\rho}_{ij}$ on Differences in Risk Factor Exposures	39
1.9	Weekly Abnormal Returns (in bps) to $\Delta\beta$ Trading Strategy	51
2.1	VAR Estimation	85
2.2	VAR Residual Correlations and Standard Deviations	91
2.3	VAR Residual Autocorrelations	92
2.4	VAR Specification Test	93
2.5	Cash-flow, Discount-rate, and Variance News for the Market Portfolio	95
2.6	Forecasting Long-Horizon Realized Variance	100
2.7	Cash-flow, Discount-rate, and Variance Betas in the Early Sample	109
2.8	Cash-flow, Discount-rate, and Variance Betas in the Modern Sample	111
2.9	Betas for Six Risk-Sorted Portfolios in the Early Sample	113
2.10	Betas for Six Risk-Sorted Portfolios in the Modern Sample	115
2.11	Betas for Option, equity, and bond portfolios	117
2.12	Asset Pricing Tests for the Early Sample	121

2.13	Asset Pricing Tests for the Modern Sample	124
2.14	Early Sample Asset Pricing Tests with Risk-sorted Portfolios	129
2.15	Modern Sample Asset Pricing Tests with Risk-sorted Portfolios . . .	132
3.1	Financial sector components in national income accounts	144
3.2	Time series summary statistics and correlations	185
3.3	Time series regressions of first differences	188
3.4	Summary Statistics for Panel Data Analysis	194
3.5	Base panel regression with time trend	199
3.6	Testing the May Day effect in the time series	201
3.7	Testing May Day effect in the cross-section	204

THIS WORK IS DEDICATED TO ARIELLA, STUART, AND SPENCER.

Acknowledgments

THE RESEARCH PRESENTED HERE was motivated by numerous conversations with the members of my dissertation committee, including John Campbell, Josh Coval and Alp Simsek, as well as significant contributions from Robin Greenwood, Jeremy Stein and Andrei Shleifer. Above all, I thank the tireless support of Brittany Lott Turley, who inspired and supported me throughout the endeavor.

The energies and skills of the professional investor and speculator...are, in fact, largely concerned, not with making superior long-term forecasts of the probable yield of an investment over its whole life, but with foreseeing changes in the conventional basis of valuation a short time ahead of the general public.

John Maynard Keynes

1

Price Comovement and Time Horizon: Fads and Fundamentals

THE INVESTMENT RISK OF A PORTFOLIO is closely connected to the comovement of its components; risk diversifies when price movements are independent but persists when changes in price are correlated. But what if prices move together over short time intervals but seem less related over long horizons? It would seem they share exposure to a fad that is unrelated to fundamental risk or profitability. In other cases, closely related assets might have prices that move together over long horizons but not over shorter intervals. This insufficient comovement masks their shared fundamental

exposures. Analyzing the returns to individual US equities, I find their correlations depend significantly on the time horizon considered. For each pair of stocks, measures of shared trading behavior versus measures of shared fundamentals are highly predictive of excess or insufficient comovement.

My empirical results employ a novel methodology in estimating how much of the measured differences in short-horizon and long-horizon correlations arise from estimation noise. This drives the statistical inference, emphasizing that these differences are too large to be circumstantial. The weekly returns to a typical pair of US stocks have a correlation of 18%, but I find the correlation of their 6-month returns are frequently 20% higher or lower than their weekly returns would suggest. Long-horizon correlations predictably decrease for stocks with similar investor trading patterns and correlations predictably increase for stocks of firms with closely related business prospects as measured by their industry affiliation or by past accounting measures.

In contrast with previous studies studying excess comovement by looking for special cases where nominal labels change but fundamental risks do not, I take the broad universe of US stocks and analyze comovement through differences in short-run and long-run correlations. The methodology could easily be employed within or across other asset classes.

Correlations are a key ingredient in asset allocation and asset pricing, and these findings have practical implications for investors. Estimates of portfolio risk should depend on the time horizon. Buy-and-hold investors may be misled if their diversification estimates are based on short-term returns. Short-horizon correlations

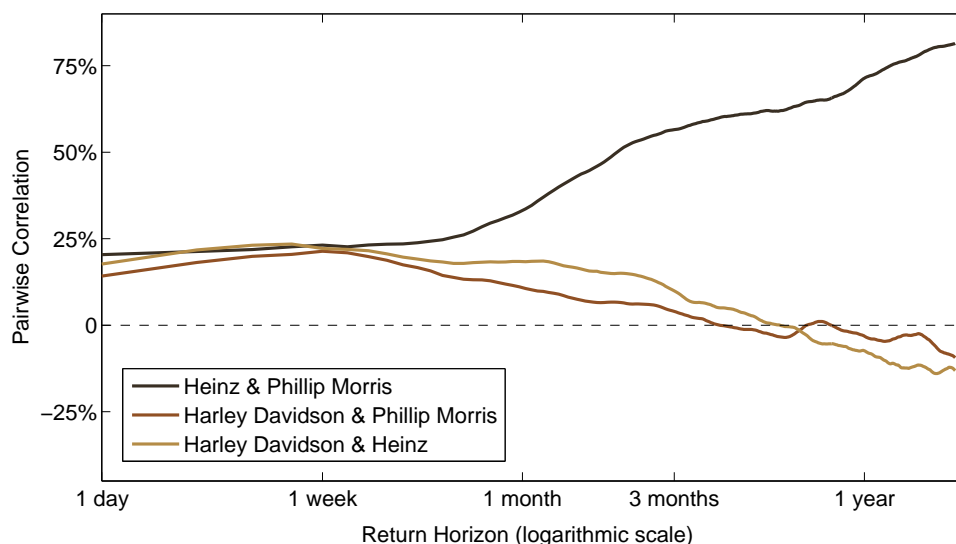


Figure 1.1: Correlations for Heinz, Phillip Morris and Harley Davidson

will be much more pertinent to an investor who rebalances frequently. Such an investor might also take advantage of the associated predictability. A simple long/short trading strategy based on a measure of fads versus fundamentals generates risk-adjusted annual excess returns of 8.4% and a Sharpe Ratio of 1.03.

As a motivating example, consider the returns to three large US stocks, Heinz, Philip Morris, and Harley Davidson. During the 1990's, all three stocks were actively traded, and their business lines were relatively stable until the turn of the century, when Philip Morris began a series of acquisitions and divestitures. Looking at their weekly returns during this decade, each pairing of the three firms has a correlation of approximately 20%. This is slightly greater than the average correlation we observe for most large cap US stocks during this period.

Now consider the long-run fundamentals shared by these stocks. Although

popular culture might lead you to connect the customers of Philip Morris' tobacco products with the stereotypical motorcyclist astride a Harley, some of the largest business lines of Philip Morris included more traditional food staple brands such as Kraft, Oscar Mayer and Jell-O. As you might expect, Philip Morris' accounting profits correlated with those of Heinz (quarterly ROE correlation of 27%), another producer of food staples, yet seem to have no relationship with those of Harley Davidson.

These relationships become increasingly apparent as the time horizon for returns lengthens and the estimated correlations differ significantly from the one-week estimates. Figure 1.1 shows how the correlation estimates change with the length of the return interval used within the decade. As the horizon increases, the correlation of the returns of Philip Morris and Heinz steadily increases to greater than 70%, while the correlations of each firm's returns with those of Harley Davidson decrease to approximately zero.

Admittedly, the examples of Heinz, Philip Morris and Harley Davidson are selected ex post from an enormous number of pairwise correlations and possible sample periods. Estimates of long-horizon correlations are noisy and the plots in Figure 1.1 could be coincidental. A more careful analysis of US stock returns between 1970 and 2010 confirms patterns of this sort are pervasive.

A number of researchers have highlighted characteristics that appear to drive excess comovement in equity returns. Barberis, Shleifer and Wurgler (2005) and Boyer (2011) consider equity index inclusion and find that the addition of a stock to major market indices causes an immediate increase in the correlation of its returns with other index constituents. Similarly, Brealey, Cooper and Kaplanis (2009) look at

changes in exchange listing due to cross-border mergers and find a stock's comovement immediately increases with securities listed in its new home market. Controlling even more strongly for differences in fundamental risk, Dabora and Froot (1999) look at companies with shares that trade on multiple exchanges and find that the prices of otherwise identical claims diverge from each other and move with other stocks listed on their respective exchanges. The empirical strategy employed in each of these papers compares comovement in a specific subset of stocks for which circumstances suggest there are no differences in fundamental risk, at least on average.

In contrast, my approach examines a broad universe of stock prices and seeks to measure the aggregate extent to which fads and fundamentals drive comovement. Instead of comparing correlations immediately before and after some event, I compare correlations made over the exact same time period where the only difference is the return increment. In this respect, there are fewer concerns about omitted risks associated with the treatment effect.

The study of excess comovement and fundamentals bears similarity to the work motivated by Shiller (1981), questioning how the aggregate stock market can be so volatile compared to the relatively stable pattern of dividends received by investors. This led to a large literature testing variance ratios over various time horizons. There are two advantages to studying correlations rather than variance ratios. First, correlations control for volatility and are less affected by time variation in market discount rates. Second, the rich cross-section of correlations allows for panel analysis, avoiding many of the econometric shortcomings associated with analyzing long-horizon returns in a limited time series.

One of the more striking empirical features of equity correlations is the fact that the historical correlations between most stocks increase as their return horizon lengthens. This stylized fact has not gone unnoticed. Campbell, Lettau, Burton and Xu (2001) study the volatility of individual equities and note how equity correlations generally declined during the 1980's and 1990's and how correlation estimates using daily returns are, on average, lower than those using monthly returns. Lo and MacKinlay (1990) study the profitability of contrarian strategies and attribute the success of this strategy to positive cross-autocorrelation. Their conclusions imply that correlations increase with time horizon. This is historically true, though I show much of this effect is due to market microstructure and becomes less prominent as trading costs have decreased.

What sort of labels might be most salient for investors' fads? Since market capitalization and relative valuations are common groupings, we might associate fads with investment styles based on size and value. This is a key prediction of Barberis and Shleifer (2003), who propose style driven investing accommodates the cognitive limitations of investors. Veldkamp (2006) derives similar predictions in a rational setting where investors generalize costly information across similar firms. My empirical results show weak evidence that firms of a similar size exhibit excess comovement, and my results do not show excess comovement in firms with similar book-to-market ratios.

Others have connected evidence of excess comovement with trading patterns by obtaining trade or position data for retail investors (Kumar and Lee, 2006) and mutual fund managers (Greenwood and Thesmar, 2011; Antón and Polk, 2010).

Given the increasing importance of index benchmarks, Greenwood (2008) looks at how index construction can lead to return patterns induced by index based trading. In this paper, I attempt to measure shared trading behavior directly by using the mechanical autocorrelations in returns caused by bid-ask bounce (Roll, 1984) or the temporary market impact of trading (Campbell, Grossman and Wang, 1993).

To measure shared fundamentals, my primary measure is the past correlation of accounting returns, measured by return on equity (ROE). I also look at common industry membership as an indicator that firms face similar demand or profitability shocks. The attempt to connect stock comovement to fundamentals builds on the work of Pindyck and Rotemberg (1993), who find most price comovement is unrelated to macroeconomic shocks and Cohen, Polk and Vuolteenaho (2009), who find the CAPM performs better when they measure betas using accounting returns rather than traditional price return betas.

The relationship between return horizon and correlation serves as a valuable measure of excess comovement in asset prices. It quantifies the economic significance of previous studies that identify an individual phenomenon driving excess comovement. By introducing measures of trading behavior and fundamentals, I can further identify the fads associated with excess comovement and the insufficient comovement associated with shared fundamentals. This is a natural framework to think about risk and portfolio construction, which yields intuition for portfolio management and asset prices.

1.1 MODELING AND MEASURING COMOVEMENT

To better understand how correlations might change with time horizon, consider what happens to the comovement of asset prices if investors are slow in incorporating new information about fundamental value and if swings in the popularity of investments affect their demand. We can contrast this with the case of no return predictability or where return predictability comes through long-term time variation in discount rates. This simple model of fads and fundamentals also suggests a prediction regarding which pairs of assets will show correlations increasing with time horizon and which pairs of assets will show decreasing correlations.

The model could apply to any sort of financial asset or portfolio of assets. The effect of time horizon on correlation is likely greatest in cases where markets are segmented or where the fundamental value is opaque. However, the notation and presentation of the model will consider the assets to be individual equity securities, in line with the empirical analysis to be presented.

MODELING FADS AND FUNDAMENTALS

Define the fundamental value of security i at time t as $P_{i,t}^*$, entitling its owner to payout $D_{i,t+1}$. Changes in log value, $\Delta p_{i,t+1}^* = \ln \frac{P_{i,t+1}^* + D_{i,t+1}}{P_{i,t}^*}$ will be a combination of the expected return and the unexpected shock,

$$\Delta p_{i,t+1}^* = E_t [\Delta p_{i,t+1}^*] + \eta_{i,t+1}. \quad (1.1)$$

Suppose that the market price may differ from this fundamental value for two reasons: first, transitory fads may cause short-run price deviations across certain groups of securities, and second, changes in fundamental value may be incorporated with a delay. This can be modeled in a simple way by defining the log return to security i as

$$r_{i,t+1} = \Delta p_{i,t+1}^* - \Delta d_{i,t+1} + \Delta f_{i,t+1} \quad (1.2)$$

where the delay in incorporating fundamentals, $\Delta d_{i,t+1}$, is governed by $\delta_d \in [0, 1)$ in

$$\Delta d_{i,t+1} = \eta_{i,t+1} - (1 - \delta_d) \sum_{k=0}^{\infty} \delta_d^k \eta_{i,t-k+1}, \quad (1.3)$$

and the fad component,

$$\Delta f_{i,t+1} = \varepsilon_{i,t+1} - \frac{1 - \delta_f}{\delta_f} \sum_{k=1}^{\infty} \delta_f^k \varepsilon_{i,t-k+1}, \quad (1.4)$$

has shocks $\varepsilon_{i,t+1}$ that decay through $\delta_f \in [0, 1)$. I will assume that $\eta_{i,t}$ and $\varepsilon_{i,t}$ are independent martingale difference sequences.

Although this implies predictability in returns, it may not be easy to recognize.

These two forces have offsetting effects on univariate tests of predictability. For example, consider an attempt to detect forecastability using the autocovariance. For simplicity, we'll assume for now that expected returns change very little (i.e.

$\text{Cov}[E_t[\Delta p_{t+1}^*], E_{t+\tau-1}[\Delta p_{t+\tau}^*]] \approx 0$)¹. The autocovariance of r_t with return $r_{t+\tau}$

¹Note that short-term variation could be driven by behavioral or rational causes, but the label "fad" will be used to categorized price movement that is transient and over very short horizons. The empirical impact of time variation in discount rates is specifically addressed in Section 1.5.

realized $\tau > 0$ periods in the future is

$$\text{Cov}[r_t, r_{t+\tau}] = \underbrace{\delta_d^\tau \left(\text{Var}[\eta_{i,t} - \Delta d_{i,t}] \right)}_{\text{momentum in fundamentals}} - \underbrace{\delta_f^\tau \left(\delta_f^{-1} \text{Var}[\Delta f_{i,t} - \varepsilon_{i,t}] \right)}_{\text{reversal in fads}}. \quad (1.5)$$

The delays in incorporating information contribute to momentum in returns (positive autocorrelation), but the transient nature of fads contribute to return reversal (negative autocorrelation). These may offset enough that it is hard for an autocorrelation or variance ratio test to reject the null hypothesis of no predictability.

Fortunately, we may be able to take advantage of variation in the way fads and fundamentals affect different assets. In the context of this model, there will be an asset j for which we can measure the effect of the fad (the correlation of $\varepsilon_{i,t}$ with $\varepsilon_{j,t}$) or delayed fundamentals (the correlation of $\eta_{i,t}$ with $\eta_{j,t}$). A temporary increase in the popularity of blue chip stocks, for example, may cause the prices of these firms to rise together even when their future earnings are unchanged and unrelated. Measures of comovement across assets could offer better information regarding the extent to which prices temporarily deviate from fundamentals.

DEFINING COMOVEMENT

To be more precise in defining comovement, I will generally refer to the short-term comovement of asset i and asset j as their contemporaneous correlation

$$\rho_{ij}(1) = \frac{\text{Cov}[r_{i,t+1}, r_{j,t+1}]}{\sqrt{\text{Var}[r_{i,t+1}] \text{Var}[r_{j,t+1}]}}. \quad (1.6)$$

The long-horizon return of asset i over H periods will be $\sum_{h=1}^H r_{i,t+h}$, so the long-term comovement of asset i and asset j is then the correlation associated with their returns with horizon length H ,

$$\rho_{ij}(H) = \frac{\text{Cov} \left[\sum_{h=1}^H r_{i,t+h}, \sum_{h=1}^H r_{j,t+h} \right]}{\sqrt{\text{Var} \left[\sum_{h=1}^H r_{i,t+h} \right] \text{Var} \left[\sum_{h=1}^H r_{j,t+h} \right]}}. \quad (1.7)$$

One advantage of measuring comovement through correlations is that it controls for changes in the variance of assets i and j in the denominator. In that sense we are focusing on their joint price behavior as opposed to factors affecting their individual volatilities. A key result comes from expanding the variance and covariance terms in the definition of long-term correlation,

$$\begin{aligned} \text{Cov} \left[\sum_{h=1}^H r_{i,t+h}, \sum_{h=1}^H r_{j,t+h} \right] &= \sum_{h=1}^H \text{Cov} [r_{i,t+h}, r_{j,t+h}] + \sum_{k \neq h} \sum_{h=1}^H \text{Cov} [r_{i,t+h}, r_{j,t+k}] \\ \text{Var} \left[\sum_{h=1}^H r_{i,t+h} \right] &= \sum_{h=1}^H \text{Var} [r_{i,t+h}] + \sum_{k \neq h} \sum_{h=1}^H \text{Cov} [r_{i,t+h}, r_{i,t+k}]. \end{aligned} \quad (1.8)$$

The assumption of no fads or delayed fundamentals means past returns do not forecast the future. This implies $\text{Cov}[r_{i,t+h}, r_{j,t}] = 0 \forall j$ and $\forall h \neq 0$, so the double summations in the equations above must equal zero. In this case

$$\rho_{ij}(H) = \rho_{ij}(1) \forall H, \quad (1.9)$$

and correlations should be the same regardless of return horizon. We might denote

the difference between long-run and short-run correlations as

$\Delta\rho_{ij} = \rho_{ij}(H) - \rho_{ij}(1)$. My null hypothesis is $\Delta\rho = 0$. As an alternative, I propose $\text{Cov}[r_{i,t+h}, r_{j,t}] \neq 0$ and is instead

$$\text{Cov}[r_{i,t+h}, r_{j,t}] = \underbrace{\rho_d^\tau \left(\text{Cov}[\eta_{i,t} - \Delta d_{i,t}, \eta_{j,t} - \Delta d_{j,t}] \right)}_{\text{shared fundamentals}} - \underbrace{\rho_f^\tau \left(\rho_f^{-1} \text{Cov}[\Delta f_{i,t} - \varepsilon_{i,t}, \Delta f_{j,t} - \varepsilon_{j,t}] \right)}_{\text{shared fads}}. \quad (1.10)$$

This will be positive when the first term is more important for a pair of firms and negative when the second term dominates. Correlations will no longer remain consistent regardless of time horizon. Instead, equation (1.8) shows how firms with similar fundamentals will have correlations that increase with time horizon and firms whose prices share exposure to fads will have correlations that decrease with time horizon.

EMPIRICAL ESTIMATION OF COMOVEMENT

In estimating the relationships of long-horizon returns can be problematic within a given sample. The sample size effectively gets smaller as the return horizon increases. For example, with a return horizon of six months, a decade of data allows for only twenty independent increments. Additionally, the long-horizon returns within a given sample will depend on the start and end dates chosen. Six month returns starting in January and June might yield different results than returns starting in April and October. We can minimize the impact of these limitations by estimating

correlations using every possible overlapping window available.

Within a given sample, a correlation for horizon length H is estimated as

$$\hat{\rho}_{ij}(H) = \frac{\sum_{h=-H}^H \left(\frac{H-h}{H}\right) \hat{c}_{ij}(h)}{\sqrt{\left(\sum_{h=-H}^H \left(\frac{H-h}{H}\right) \hat{c}_{ii}(h)\right) \left(\sum_{h=-H}^H \left(\frac{H-h}{H}\right) \hat{c}_{jj}(h)\right)}}. \quad (1.11)$$

The empirical cross-autocovariance $\hat{c}_{ij}(h)$ measures the relationship between r_i and r_j 's realizations of h periods in the future,

$$\hat{c}_{ij}(h) = \frac{1}{H-r} \sum (r_{i,t} - \bar{r}_i) (r_{j,t+r} - \bar{r}_j). \quad (1.12)$$

Estimating long-run correlations using (1.11) is equivalent to averaging the correlation estimates for returns of horizon length H using all possible windows. Suggestively, this is also identical to the correlation resulting from Newey and West's (1987) estimator of the long-run covariance of a time series. The fundamental risk in a financial time series is closely related to the concept of long-run variance, which continues to be a major topic of research in time series econometrics.

THE PRICE IMPACT OF TRADING BEHAVIOR

To identify the sorts of firms whose prices are driven by shared trading behavior rather than fundamentals, we could propose characteristics that might be overly salient to investors and test to see if they predict negative values for $\Delta\rho_{ij}$. For example, if investment styles are indicative of non-fundamental related trading they would show negative coefficients in a regression.

To capture trading behavior more directly, we can try to measure which assets tend

to be contemporaneously bought and sold. The simple model above would predict that assets with a greater degree of shared trading behavior will exhibit more values for $\Delta\rho_{ij}$. While it might seem difficult to observe data on who is initiating transactions, I will show how shared trading behavior can be inferred by looking at correlations in bid-ask bounce.

Consider Roll's (1984) model of the effective bid-ask spread. He notes that the closing price recorded for a security can be affected by whether the last trade was driven by a purchase or a sale. This price differential can be interpreted as the literal bid-ask spread paid by buyers and sellers who initiate trades with market makers, or this could be a more modern concept of temporary price impact as the intensity of buying or selling pressure affects liquidity provision.

Suppose that an average sized buyer must pay $p_{i,t} + b_i$, and sellers of an average quantity receive $p_{i,t} - b_i$. Hence b_i can be thought of as the temporary market impact of trading. Any permanent impact from information in trades is captured by updates in $p_{i,t}$. The observed return is then a combination of the price change and the transitory market impact of purchases (indicated by binary variable $\eta_{i,t} = 1$) or sales (when $\eta_{i,t} = -1$). The observed return ($\tilde{r}_{i,t+1}$) can be expressed as the log return ($r_{i,t+1} = p_{i,t+1} - p_{i,t}$) plus the market impact

$$\tilde{r}_{i,t+1} = r_{i,t+1} + b_i \left(\eta_{i,t+1} - \eta_{i,t} \right). \quad (1.13)$$

Let's assume that purchases and sales are equally likely and are independent each period and the null hypothesis that past price changes are not predictive of the future.

The effect of this trading on the autocovariance sequence for returns will be

$$\begin{aligned}
\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{i,t}] &= \text{Var} [p_{i,t+1} - p_{i,t+1}] + b_i^2 \\
\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{i,t+1}] &= -b_i^2 \\
\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{i,t+k}] &= 0 \quad \forall k > 1.
\end{aligned} \tag{1.14}$$

This is precisely what motivated Roll's estimate of the effective bid-ask spread:

$$b_i = -\sqrt{\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{i,t+1}]}.$$
(1.15)

And what if the buying pressure is correlated across firms? Suppose that investors tend to buy and sell asset i and asset j at the same time, so that $v_{ij} = \mathbb{E} [\eta_{i,t}, \eta_{j,t}] \neq 0$. We would observe $v_{ij} > 0$ if the trading behavior is similar and $v_{ij} < 0$ if investors tend to buy one while selling the other. Intuitively, we can write v_{ij} as a simple function of the probability that securities are both exposed to common trading behavior,

$$v_{ij} = 2 \times \left(\Pr [\eta_{i,t} = \eta_{j,t}] - 0.5 \right).$$
(1.16)

This is the proposed measure of common trading behavior. Just as we can measure the effective bid-ask from the autocovariances, we can estimate common trading behavior from the cross-autocovariances. Under the same assumptions as above, they

will be

$$\begin{aligned}
\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{j,t}] &= \text{Cov} [r_{i,t+1}, r_{j,t+1}] + 2v_{ij}b_ib_j \\
\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{j,t+1}] &= -v_{ij}b_ib_j \\
\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{j,t+k}] &= 0 \quad \forall k > 1.
\end{aligned} \tag{1.17}$$

From this, I empirically estimate this measure v_{ij} of how trading behavior connects two stocks through

$$v_{ij} = -\frac{\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{j,t+1}] + \text{Cov} [\tilde{r}_{i,t+1}, \tilde{r}_{j,t}]}{2\sqrt{\text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{j,t+1}] \text{Cov} [\tilde{r}_{i,t}, \tilde{r}_{j,t+1}]} }. \tag{1.18}$$

1.2 SHORT-RUN AND LONG-RUN COMOVEMENT IN US EQUITIES

DATA SOURCES AND VARIABLE CONSTRUCTION

To estimate the comovement of US equity prices, I use four decades of weekly total returns from The Center for Research in Security Prices² (CRSP), covering the forty years from 1970 to 2009, and each decade is considered a subsample. To ensure the analysis focuses on the most liquid securities, I select the 2,000 largest issues by market cap as determined immediately prior to the start of each decade. The weekly log returns are measured using Tuesday's closing prices and include any distributions received. For the most recent decade spanning 2000-2009, the universe consists of the largest 2,000 firms measured by their market cap on December 31st, 1999, and

²Center for Research in Security Prices. ©2011 Booth School of Business, The University of Chicago. Used with permission. All rights reserved. www.crsp.chicagobooth.edu

the first weekly return is measured from January 4th to January 11th, 2000. Only publicly traded common stock of US incorporated firms are considered (CRSP share codes 10 and 11).

Within each decade, short-run and long-run correlations are calculated for every pair of firms, where the short run is defined as one week and the long run is defined as half of a year. Short-run correlations of weekly returns, $\hat{\rho}(1)$ are calculated as in (1.6). The long-run correlation calculation uses the formula in (1.7) where $H = 26$ weeks, generating $\hat{\rho}(26)$. The difference between the two yields $\Delta\hat{\rho}$.

To minimize any bias related to survivorship, long-run correlations are calculated whenever possible, even when two firms coexist for only a small portion of the decade. The minimum possible number of observations to calculate $\hat{\rho}(26)$ is approximately one year. The trade-off for reducing this bias is sampling variance, as the long-run variance in those cases is exceptionally noisy. In practice, requiring a longer minimum history decreases the sample size and affects the results very little, so I make this criterion as permissive as possible.

We can be reasonably comfortable that the results of the empirical analysis are not driven by the anomalous behavior of illiquid firms since the universe consists of the largest 2,000 securities by market capitalization and the shortest time interval considered is one week. The mean difference between short-run and long-run correlation increases when using smaller firms and shorter time horizons, and there is also a slight increase in the predictability of this difference, but these results are excluded as they would be open to criticism that they are affected to a larger extent by

Table 1.1: Data Coverage for Correlation Estimates

This table reports the data availability for the estimated return correlations. The return series considered are log returns calculated from the CRSP total return data, and the minimum unit of measurement is one week, corresponding to returns from Tuesday to Tuesday. The unique correlation estimates correspond to the upper triangle of the matrix of correlation coefficients, excluding the diagonal.

	Decade				Total
	1970's	1980's	1990's	2000's	
max possible pairs	1,999,000	1,999,000	1,999,000	1,999,000	7,996,000
pairs w/ min # returns	1,872,110	1,811,088	1,828,826	1,632,793	7,144,817

stale prices or other liquidity related issues.

SUMMARIZING THE CORRELATIONS OVER LONG AND SHORT HORIZONS

Summary statistics for the correlation estimates are shown in Table 1.2. The sample size of 2,000 firms will generate slightly less than two million unique correlation estimates each decade. The first panel shows the effect of attrition on data coverage. You can see that correlations can be calculated for more than 90% of all possible pairs of firms except in the most recent decade where the ten-year period begins in the year 2000, at the peak of the Internet frenzy. Acquisitions and failures cause an atypical number of firms to disappear during the first 12 months of this subsample.

For the four decades considered, the short-run correlation, $\hat{\rho}_{ij}(1)$, averages 18.4%, with a standard deviation of 11.4%. In contrast, long-run correlations are much higher, with a full sample average of 30.0% and standard deviation of 27.0%. The difference between the two, $\hat{\rho}(H) - \hat{\rho}_{ij}(1)$, averages 11.6%. The difference decreases

Table 1.2: Summary Statistics for Correlation Estimates

This table reports the data availability and summary statistics for the estimated return correlations. The return series considered are log returns calculated from the CRSP total return data, and the minimum unit of measurement is one week, corresponding to returns from Tuesday to Tuesday. The short run correlation measures, $\hat{\rho}_{ij}(1)$, are therefore associated with a one week horizon. In the data panel measuring coverage by unique correlation pairs, the unique correlation estimates correspond to the upper triangle of the matrix of correlation coefficients, excluding the diagonal.

<i>short-horizon correlation</i>		Decade				Full Sample
		1970's	1980's	1990's	2000's	
$\hat{\rho}_{ij}(1)$	mean	22.81	19.20	13.01	18.31	18.36
	std dev	9.12	10.55	9.31	13.90	11.36
	5 %ile	8.55	2.59	-0.90	-5.16	0.53
	median	22.57	18.96	12.43	18.53	18.18
	95 %ile	37.84	36.74	28.88	40.73	36.96
<i>long-horizon correlation</i>		Decade				Full Sample
		1970's	1980's	1990's	2000's	
$\hat{\rho}_{ij}(26)$	mean	45.12	30.03	22.77	20.72	30.00
	std dev	19.98	24.79	24.84	30.65	26.94
	5 %ile	10.52	-15.72	-18.44	-37.30	-18.62
	median	46.69	32.80	23.06	24.26	32.87
	95 %ile	74.95	65.68	63.59	64.71	68.81
<i>correlation difference</i>		Decade				Full Sample
		1970's	1980's	1990's	2000's	
$\Delta\hat{\rho}_{ij}$	mean	22.31	10.83	9.76	2.42	11.64
	std dev	18.36	22.21	22.73	26.23	23.52
	5 %ile	-9.37	-28.15	-27.77	-45.68	-29.86
	median	23.70	12.30	10.01	4.91	13.55
	95 %ile	49.47	44.63	46.79	40.93	46.43

over time, with an average difference of 22.3% in the 1970's decreasing to a difference of only 2.4% in the most recent decade.

By definition, there are upper and lower bounds on the possible observed correlations. In practice, the estimated short-run correlations are nearly always positive, with less than 5% of the estimated values being less than zero. However, there is much more variation in the long horizon correlation estimates. Even though the average long-run correlation is nearly twice as large, a little less than a third of the estimates are less than zero.

While the standard deviations and percentiles shown in Table 1.2 make it tempting to conclude that there is a larger degree of cross-sectional variation in correlations measured over long horizons, it is important to note that the short-run correlations are estimated much more precisely. Even under the null hypothesis where the true correlation does not depend on the time horizon, the empirical long-run correlations will show more variation due to the fact that they are estimated using far fewer independent observations. We cannot yet draw conclusions about the distribution of the true long-run correlations. The full sample standard deviation of 27.0% reflects both the dispersion of correlations in the population as well as the measurement error. The subsequent section will present a method for quantifying the effect of measurement error in the long run estimates.

1.3 A REGRESSION METHODOLOGY FOR CORRELATIONS

REGRESSING EXPLANATORY VARIABLES ON THE CORRELATION DIFFERENCES

To test the null hypothesis in (1.9) against the alternative, I propose running a regression of the difference in long-run and short-run correlation on candidate explanatory variables for each pair of firms. Negative values for this difference in correlations correspond to excess comovement, indicating the pair of stocks has a higher correlation in the short run than can be justified by their long-run returns. Positive values are indicative of insufficient comovement, as the short-run returns do not seem to capture the comovement observed over longer horizons.

Given explanatory variables corresponding to each pair of firms (i, j) whose shared characteristics constitute vector Z_{ij} (including a constant term), the coefficient vector β is estimated from the linear regression

$$\Delta \hat{\rho}_{ij} = \beta Z_{ij} + e_{ij}. \quad (1.19)$$

Under the null hypothesis, every element of β , including the constant, is equal to zero.

Calculating the standard errors for $\hat{\beta}$ requires special attention, since these errors are not independent across pairs of firms. The traditional standard errors estimated using an OLS regression to estimate (1.19) will be far too small. What appears to be a large cross-sectional sample is effectively smaller since much of the variation in stock returns is driven by common factors. Even worse, all stocks likely have a positive loading on a single factor, the market. If none of the residuals are independent,

traditional techniques to handle correlated residuals in a cross-sectional regression, like clustering standard errors, will offer little help.

A RESHUFFLING TECHNIQUE FOR STATISTICAL INFERENCE

The problem would benefit from a new approach. Note that under the null hypothesis, this error term e_{ij} is equal to the estimation error between the true long-horizon correlation and whatever empirical estimate results from the particular sample used. We can call this estimation error

$$\varepsilon_{ij} = \Delta \hat{\rho}_{ij} - \Delta \rho_{ij}(H), \quad (1.20)$$

and note that $e_{ij} = \varepsilon_{ij}$, under the null hypothesis.

Fortunately we can take advantage of some properties of the null hypothesis. In particular, the assumption of no predictability suggests that the error terms in (1.20) result from the purely coincidental estimation noise of past returns appearing to predict the future.

Therefore, the historical ordering of the weekly returns makes no difference. We just need to preserve the contemporaneous return structure. In fact, if we randomly reshuffle the historical ordering of the weeks and recalculate the long-run correlations, we would generate an independent draw of error terms with the same statistical properties.

This is effectively what I propose as a robust, non-parametric method for calculating standard errors. With new long-term correlation estimates from each reshuffling of the weekly returns, we find the distribution of β under the null by

repeatedly rerunning the regression in (1.19). Then we can compare our $\hat{\beta}$ estimate to the distribution of estimates generated from the reshuffled data. We can now test the hypothesis that $\hat{\beta} = 0$ properly accounting for the strong dependence across our observations.

The reshuffling technique also makes it possible to revisit the variation in the estimated long-horizon correlations. The observed differences in long-horizon and short-horizon correlations are due to both the variation expected from sampling noise as well as the true dispersion in correlation values. A casual glance at the magnitudes might lead someone to prematurely reject the null hypothesis based solely on the large variation in Table 1.2. The two panels in Figure 1.2 plot a histogram of the cross-sectional variation in the estimated $\Delta\hat{\rho}_{ij}$ against the density function of the sampling error expected under the null hypothesis for the earliest and the most recent decade.

Figure 1.2 also graphically emphasizes the difference between the previously documented observation that correlations seem to increase with time horizon on average (Campbell et al., 2001) and the claim that some correlations increase with horizon and some decrease. By inspection, the estimated long-horizon correlations are significantly higher than what would be expected under the null hypothesis for the 1970's, though the significance of the difference is less obvious in the 2000's. This paper will show empirical analysis suggesting that the earlier difference in mean correlation differences can be largely attributed to microstructure noise from the bid-ask spread.

Setting aside differences in the mean, the dispersion in the reshuffled values is

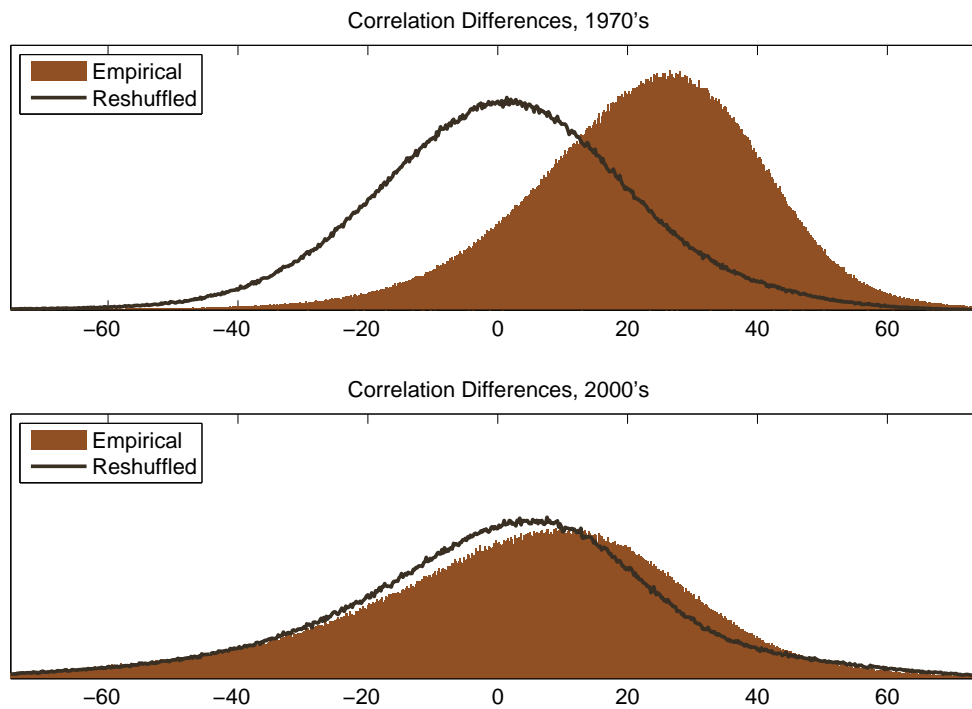


Figure 1.2: Comparing Empirical and Reshuffled Correlation Differences

quite high, suggesting that we cannot immediately rule out the possibility that large cross-sectional differences in correlation estimates for different time horizons are simply sampling error. A more careful analysis will show evidence that correlations will predictably increase or decrease as the return horizon lengthens.

1.4 EXPLAINING EMPIRICAL CORRELATIONS

DATA DESCRIPTION OF EXPLANATORY VARIABLES

All of the explanatory variables that form the elements of the Z_{ij} vector of explanatory variables in estimating (1.19) are calculated using data available prior to each decade. I group them by factors ostensibly related to investment behavior and factors that are indicative of shared fundamental risks.

I estimate shared trading behavior by calculating the correlations in bid-ask bounce, v_{ij} , as defined in (1.18). Log weekly returns are used to estimate v_{ij} using a two year window prior to the start of the decade. The effective bid/ask spread, used in the denominator of the definition of shared trading behavior is shrunk toward the median value estimated across all securities, which prevents a negative implied spread in most cases. To further control for large outliers that may be driven by a very small denominator, or by estimation error in the numerator, the final values of v_{ij} are all shrunk toward zero.

Somewhat surprisingly, Table 1.4 shows that, on average, firms do not tend to be bought and sold together for the first two decades in the sample. This might be indicative that the trading behavior tended to reflect investors shifting investments across stocks rather than a pattern of broad net inflows or outflows in the equity market. For the more recent two decades, however, the mean coefficient is much closer to zero and shows no particular propensity for stocks to be bought or sold together, though this varies significantly across pairs of stocks.

Table 1.3: Summary Statistics for Primary Explanatory Variables

This table reports the data availability and summary statistics for the explanatory variables used in the regression analysis. The summary of unique correlation pairs represent the upper triangle of the correlation matrix, excluding the own correlations on the diagonal. The shared trading behavior is an estimate of the propensity of buyers and sellers of firms to have correlations in the temporary market impact they cause, as measured through temporary components in autocorrelations. The primary variable representing fundamental correlation is the correlation of firms return on equity, as derived from quarterly accounting data from Compustat. Dummy variables capture shared characteristics related to primary trading exchange and market cap quintiles, using data from CRSP, and the book equity (BE) and GICS industry data are obtained from the linked CRSP-Compustat database.

Data Availability					
	Decade				Full Sample
	1970's	1980's	1990's	2000's	
pairs w/ min # returns	1,872,110	1,811,088	1,828,826	1,632,793	7,144,817
with $v_{i,j}$ values	1,280,586	904,756	1,482,243	1,019,096	4,686,681
with $\text{Corr}[ROE_i, ROE_j]$	204,757	1,320,533	963,477	677,598	3,166,365
Summary Statistics					
v_{ij}	Decade				Full Sample
	1970's	1980's	1990's	2000's	
mean	-0.45	-1.22	0.02	-0.14	-0.38
std dev	1.15	1.32	0.94	1.06	1.20
5 %ile	-2.38	-3.16	-1.62	-1.97	-2.44
median	-0.42	-1.42	0.11	-0.08	-0.27
95 %ile	1.33	1.30	1.37	1.46	1.37
$\text{Corr}[ROE_i, ROE_j]$	Decade				Full Sample
	1970's	1980's	1990's	2000's	
mean	0.12	0.08	0.01	0.02	0.05
std dev	0.49	0.35	0.30	0.29	0.33
5 %ile	-0.72	-0.50	-0.48	-0.46	-0.50
median	0.16	0.08	0.01	0.01	0.04
95 %ile	0.84	0.64	0.52	0.51	0.62

My primary measure to estimate fundamental correlation is the correlation of firms' return on equity. ROE values are constructed from Compustat data, defined as the ratio of earnings per share (Compustat item: epspiq) divided by common equity per share (Compustat item: ceqq). This value is censored at -90% and +100% and then converted to a log return. Annual Compustat data is used to supplement where quarterly data is not available. Correlations in this ROE series are calculated for each pair of firms over the prior 10 years, excluding the quarter immediately prior to the beginning of the decade, since this data is typically not released until January or later. I set a minimum requirement of 4 years of accounting data to estimate a valid correlation. As can be seen in the coverage statistics in Table 2, lack of Compustat data tends to be the most restrictive data requirement, especially near the beginning of the sample when only a few hundred firms have accounting data available. This does not have a substantive effect on the regression results, but I will run a regression specification that excludes $\text{Corr}[\text{ROE}_i, \text{ROE}_j]$ to take advantage of the larger data set.

Market cap and exchange information all come from CRSP, and the book equity and GICS industry assignments are all taken from the CRSP-Compustat linked database. The construction of the book equity / market equity (BE/ME) variable mirrors that described by Fama and French (1992). Each decade, the 2000 firms in the universe are matched to their assigned to BE/ME quintiles relative to the CRSP universe of firms. I do not use the CRSP universe for market cap quintile assignments, since my sample of the 2,000 largest firms only represents the largest quintiles. Instead, I create market cap quintiles specific to this sample using market cap data from the December previous to the start of each decade.

This information allows for the construction of the dummy variables shown in Table 1.4. They correspond to pairs of firms being listed on the same exchange, sharing the same size quintile, being assigned the same GICS industry, etc. As usual, the dummy variables equal 1 for each pairwise observation where the criteria are met. The classifications of sharing the same GICS sector, industry or subindustry are not exclusive of each other, so a pair of firms in the same subindustry will necessarily also be in the same industry and sector. The occurrence of firms in the same subindustry is the rarest of the dummy variables, occurring in about 1.7% of the unique firm pairs, but will be shown to have a strong effect even after controlling for industry and sector.

1.4.1 REGRESSING EXPLANATORY VARIABLES ON $\Delta\hat{p}$

Following the methods described in section 1.3, I estimate regression coefficients for each decade subsample via least squares and use the reshuffling technique to calculate standard errors. The regression estimates for regressions of $\Delta\hat{p}$ on various explanatory variables are combined (assuming independent subsamples) and reported in Table 1.5.

The first regression specification includes no explanatory variables other than constant terms. While these regression coefficients are going to reflect the simple means previously noted in the summary statistics, the reshuffling methodology help us better understand the significance of these results. We can see that even across almost 2 million observations per decade, the common factors driving returns can generate standard errors in the average difference in long-run and short-run correlations of about 3%. The fact that long-horizon correlations average 2.42%

Table 1.4: Summary Statistics for Dummy Variables

This table reports the data availability and summary statistics for dummy variables used as explanatory variables in the regression analysis. They characteristics related to primary trading exchange and market cap quintiles use data from CRSP, and the book equity (BE) and GICS industry data are obtained from the linked CRSP-Compustat database.

Data Availability					
	Decade				Full Sample
	1970's	1980's	1990's	2000's	
pairs w/ min # returns	1,872,110	1,811,088	1,828,826	1,632,793	7,144,817
with GICS industry	686,162	1,198,845	1,771,901	1,611,183	5,268,091
with BE/ME values	1,212,759	1,512,444	1,567,333	1,167,604	5,460,140

Frequency					
	Decade				Full Sample
	1970's	1980's	1990's	2000's	
same exchange	53.2%	47.7%	44.4%	49.6%	48.7%
same size quintile	20.0%	20.0%	20.0%	20.0%	20.0%
same BE/ME quintile	20.3%	21.6%	21.8%	30.9%	23.4%
same sector	15.2%	13.1%	12.8%	15.7%	14.1%
same industry	2.7%	3.1%	3.2%	2.8%	3.0%
same subindustry	1.6%	1.8%	1.7%	1.7%	1.7%

higher than short-horizon correlations in the most recent decade is well within the range of differences we might randomly observe. The differences in earlier decades, as large as 22% during the 1970's, cannot be explained by estimation error.

The second regression specification includes the two primary explanatory variables reflecting shared trading behavior (v_{ij}) and shared fundamentals ($\text{Corr}[ROE_i, ROE_j]$). Both of these variables are highly significant in explaining the effect of return horizon on correlations. As expected, common trading behavior is indicative of temporary price comovement, as indicated by the negative coefficient. Firms that have a higher probability of being bought or sold together have higher short-horizon correlations but lower correlations over long horizons. The variable measuring shared fundamentals generates a positive regression coefficient and the opposite effect of trading behavior. Firms with highly similar fundamental exposures tend to have lower short-horizon correlations relative to long horizons, suggesting insufficient comovement.

The third regression specification adds the dummy variables indicating firms are traded on the same exchange, and in similar size or valuation categories, or belong to the same GICS industry categories. Trading on the same exchange is indicative of excess comovement, consistent with the international evidence that exchange listings matter. Considering the three principal exchanges on which these stocks are listed (NYSE, AMEX, and NASDAQ), more than 1% of stock price variation is associated with temporary comovement with other stocks on the same exchange. As is true with all the explanatory variables considered, the exchange listing may not be the causal force driving excess comovement, but it is predictive.

Table 1.5: Cross-Sectional Regressions of Correlation Difference, $\Delta\hat{\rho}_{ij}$

In the regressions below, the dependent variable is the difference between long run and short run correlation ($\Delta\hat{\rho}_{ij}$). All of the explanatory variables are dummy variables except for Shared Trading Behavior (v_{xy}) and Shared Fundamentals ($\text{Corr}[ROE_i, ROE_j]$). The reported coefficients are from combining cross-sectional regressions for each decade, and standard errors, reported in parentheses below the regression coefficients, use the reshuffling methodology described in section 1.3 for each cross-section and assume the subsamples are independent. Statistical significance of the coefficient relative to the null hypothesis of zero is denoted using asterisks, where * indicates significance at the 5% level and ** indicates significance at the 1% level.

	(1)	(2)	(3)	(4)
1970's Decade Dummy	22.31** (3.86)	18.20** (4.22)	20.02** (3.99)	19.36** (3.97)
1980's Decade Dummy	10.83** (3.31)	9.65** (3.37)	10.85** (3.57)	10.27** (3.51)
1990's Decade Dummy	9.76** (2.68)	9.01** (2.81)	9.31** (2.65)	9.30** (2.75)
2000's Decade Dummy	2.42 (3.18)	3.33 (3.62)	2.15 (3.58)	2.45 (3.76)
Shared Trading Behavior (v_{ij})		-0.82** (0.13)		-0.74** (0.13)
Same Exchange			-1.34** (0.48)	-1.85** (0.60)
Same Size Quintile			-0.43* (0.17)	-0.93** (0.28)
Same Be/ME Decile			0.58* (0.24)	0.71** (0.19)
Shared Fundamentals ($\text{Corr}[ROE_i, ROE_j]$)		1.27** (0.23)		0.98** (0.22)
Same Sector			4.49** (0.43)	4.92** (0.43)
Same Industry			1.34** (0.47)	0.26 (0.63)
Same Subindustry			2.47** (0.58)	3.31** (0.93)
Observations	7,144,817	2,208,662	4,324,466	1,946,156

The dummy variable indicating firms are in the same size quintile also has the expected sign. Prices of firms with similar market caps seem to move together over short horizons much more than over longer return horizons. On the other hand, the same logic would suggest a negative regression coefficient on the dummy variable indicating firms are in the same BE/ME quintile, but this is not the case. The coefficient on this variable is positive. A closer examination of excess comovement across subsamples and controlling for autocorrelations from market microstructure suggests the value results are not robust and the size effect is primarily driven by excess comovement in the firms at the smaller range of this sample.

The variables indicating firms share the same sector, industry or subindustry all show large positive coefficients. As with the measure of shared fundamentals that looks at correlations in profitability, these variables seem to indicate firms with similar factors driving their profitability show insufficient price comovement over short horizons. For firms in the same subindustry, the correlation of their 6-month returns will, on average, be 8.3% higher than the correlation of their weekly returns. This is one of the strongest statistical results, though it's not without precedent. Cohen and Frazzini (2008) and Moskowitz and Grinblatt (1999) show evidence of evidence of positive momentum across connected firms, which would cause their correlations to increase with the time horizon.

The fourth regression specification includes all explanatory variables. This serves as a check that each makes an independent contribution. There is a slight decrease in the coefficients on the main variables measuring shared trading behavior and shared fundamentals, but they remain highly significant.

Interestingly, the coefficients on the other variables intended to capture labels that might be salient to investors all increase. The coefficient on firms that share the same size quintile almost doubles, indicating that it might be more prominent conditioned on the other explanatory variables than it is when measured in isolation.

The variables intended to capture common exposures to fundamental risks all remain significant predictors of insufficient short-run comovement with the exception of the dummy variable for firms sharing the same industry. This is actually an artifact of this measure being so similar to the subindustry dummy variable that the coefficient shifts from one to the other.

The general conclusions from the empirical results are broadly consistent across regression specifications. They provide evidence in favor of the hypothesis that short-run comovement is different than long-run comovement, and that excess and insufficient comovement can be predicted by measures of shared trading behavior and exposures to shared fundamentals.

ROBUSTNESS

The key results in Table 1.5 are robust across a variety of alternative estimation approaches. However, there are two critiques that deserve special attention, which I'll call the "micro explanation" and the "macro explanation." The micro explanation would assert that the correlation differences are the result of bid-ask spreads and similar effects in market microstructure, and the macro explanation would assert that correlation differences are simply a manifestation of predictability in well-known risk factors.

Just as the bid-ask bounce can be used to estimate trade-driven price behavior, serial correlation from market microstructure can also affect correlations. This is clear from the effects derived in (1.14) and (1.17). In general, long-run correlations will appear mechanically higher than short-run correlations simply because the temporary price impact of trading constitutes a much smaller fraction of total price movement in long-horizon returns relative to short-horizon returns. Since this effect will be larger for stocks that are less liquid, the regression analysis might mistakenly associate measures correlated with liquidity as indicators of insufficient comovement.

To show this is not the source of the results in Table 1.5, I construct a measure that adjusts the difference between long and short-horizon correlations that excludes the first order autocorrelation and cross-autocorrelation terms that could be affected by the impact of trading on closing prices. I label this variable $\Delta\dot{\rho}_{ij}$. These excluded first order autocorrelations would also contain a large degree of information about excess comovement, so it is important to recognize that assuming them to be zero may be a useful robustness check, but it biases all results in favor of the null hypothesis.

Table 1.6 reports summary statistics for $\Delta\dot{\rho}_{ij}$. Comparing these microstructure adjusted estimates to the original summary statistics reported in Table 1.2. The most striking difference is that the mean short-run correlation is much closer to the mean long-run correlation. This suggests that the lower comovement in the short run is driven, in a large part, by the idiosyncratic price impact from trading that immediately reverses in the subsequent period. This is in line with the predicted effect of market microstructure.

Not surprisingly, the microstructure adjustments become less significant over

Table 1.6: Microstructure Robust Correlation Differences

This table reports summary statistics for the microstructure-robust correlation differences, $\Delta\dot{\rho}_{ij}$, where the autocorrelation terms in defining the long run correlation are assumed to be zero. The calculations are otherwise identical to those described for $\Delta\rho_{ij}$.

		Decade				Full Sample
		1970's	1980's	1990's	2000's	
$\Delta\dot{\rho}_{ij}$	mean	9.30	1.71	5.01	1.74	4.55
	std dev	16.13	37.68	19.14	32.59	27.83
	5 %ile	-16.74	-34.49	-25.64	-45.47	-30.07
	median	9.75	3.03	5.06	4.67	5.96
	95 %ile	33.92	35.54	35.53	38.10	35.60

time, which is likely a result of increased liquidity and tighter bid-ask spreads. The dispersion of the difference remains high on average and over time, suggesting that the return horizon may have a large effect on individual correlations, even when the difference is only slightly positive in the cross-section.

To check the robustness of the regression results directly, I run the previous regressions on $\Delta\dot{\rho}$, the difference in long-term and short-term correlations that have been adjusted for microstructure. These regression results are shown in Table 1.7.

The most noticeable differences are in the unconditional averages, as seen in the first regression specification with no other explanatory variables. As was observed in the summary statistics, the differences all decrease. Looking at the statistical significance only the 9.3% average difference in the 1970's remains statistically different from zero at the 5% confidence level. This is consistent with the idea that a great degree of the insufficient comovement we observed was an artifact of temporary impact of trades on closing prices.

Table 1.7: Regressions Adjusted for Microstructure Effects, $\Delta\dot{\rho}_{ij}$

In the regressions shown, the dependent variable is the difference between long run and short run correlation, after adjusting for the first order autocorrelation that is likely caused by bid-ask bounce and other microstructure effects, yielding $(\Delta\dot{\rho}_{ij})$. All of the explanatory variables are dummy variables except for Shared Trading Behavior (v_{xy}) and Shared Fundamentals ($\text{Corr}[ROE_i, ROE_j]$). The reported coefficients are from combining cross-sectional regressions for each decade, and standard errors, reported in parentheses below the regression coefficients, use the reshuffling methodology described in section 1.3 for each cross-section and assume the subsamples are independent. Statistical significance of the coefficient relative to the null hypothesis of zero is denoted using asterisks, where * indicates significance at the 5% level and ** indicates significance at the 1% level.

	(1)	(2)	(3)	(4)
1970's Decade Dummy	9.30* (3.86)	6.70 (4.22)	6.95 (3.99)	6.62 (3.97)
1980's Decade Dummy	1.71 (3.31)	1.03 (3.37)	2.33 (3.57)	1.97 (3.51)
1990's Decade Dummy	5.01 (2.68)	4.95 (2.81)	4.61 (2.65)	4.72 (2.75)
2000's Decade Dummy	1.74 (3.18)	4.97 (3.62)	3.94 (3.58)	5.22 (3.76)
Shared Trading Behavior (v_{ij})		-0.24 (0.13)		-0.21 (0.13)
Same Exchange			-0.70 (0.48)	-0.89 (0.60)
Same Size Quintile			-0.07 (0.17)	-0.13 (0.28)
Same Be/ME Decile			0.13 (0.24)	0.26 (0.19)
Shared Fundamentals ($\text{Corr}[ROE_i, ROE_j]$)		0.94** (0.23)		0.56* (0.22)
Same Sector			1.61** (0.43)	1.83* (0.43)
Same Industry			0.90 (0.47)	0.04 (0.63)
Same Subindustry			1.20* (0.58)	1.50 (0.93)
Observations	7,144,817	2,208,662	4,324,466	1,946,156

Although none of the explanatory variables identified as significant in the prior regression change drastically, most of their effects are more muted. For example, in the second regression specification the coefficient on the shared trading behavior variable previously had a coefficient of -0.82 and a t-statistic of -6.4, but this now drops to a coefficient of -0.24 and an associated t-statistic of -1.83. It might be that much of the temporary impact captured by this variable corrects itself in the subsequent week, which is excluded in the calculation of $\Delta \dot{p}$, or it may be that the shared trading behavior variable also proxies for liquidity.

The other main explanatory variable, measuring correlation in shared fundamentals, sees a much more moderate decrease in magnitude after adjusting for microstructure and also remains highly statistically significant. Its coefficient drops from 1.27 to 0.94.

In the fourth regression specification on Table 1.7 where all explanatory variables are included, the coefficients are generally smaller than they were in Table 1.5. The only dummy variable that could be considered statistically different from zero with greater than 95% confidence is the measure of firms being in the same GICS sector.

The assumption that long-horizon and short-horizon correlations should be equivalent comes from equation (1.8) where past returns are assumed not to predict the future. No arbitrage assumptions in asset pricing theory suggest that this should be true for conditional moments, but not necessarily true for unconditional measures of volatility and correlation. Cochrane (1991) emphasizes this point, showing how unconditional return predictability does not reject rational pricing models outright and are exactly what we could expect to see in macroeconomic models where

discount rates vary over time due to changing growth prospects or risk preferences.

The same principle holds true in our analysis. Our null hypothesis would be rejected by a broad class of models that generate time variation in the price of equity risk. Let's consider what we would expect to see in a standard model of this type. In a one-factor model where the expected returns to stocks are driven by their exposures to the aggregate stock market, time variation in expected market returns would imply that some of the short-horizon price correlation between stocks is driven by their common exposure to changes in aggregate return expectations. This common component of comovement becomes less prominent as time horizons increase. We would then expect that long-horizon correlations across all firms should, on average, be lower than short-horizon correlations. Instead, the data shows the opposite.

Additionally, we can speculate how aggregate market predictability might explain cross-sectional variation in $\Delta\rho$. Pairs of firms with large differences in their betas to priced risk factors should have lower short-run correlations relative to their long-run correlations, while firms with similar exposures should show less of a difference. If we include the absolute value of their beta differences in our regressions, we should get a positive coefficient.

I test this hypothesis by estimating firm betas for the three factor model of Fama and French (1992) prior to each decade. With firm-level coefficients for the market portfolio β_{MKT} , for the size spread portfolio, β_{SMB} , and for the value spread portfolio, β_{HML} . I calculate the absolute value of the difference in their estimated betas. These are considered as an additional explanatory variable in the cross sectional regressions of the differences in long-horizon and short-horizon correlations adjusted for

Table 1.8: Regressions of $\Delta \dot{\rho}_{ij}$ on Differences in Risk Factor Exposures

In the regressions below, the variables labeled as the $|\beta_{i,XYZ} - \beta_{j,XYZ}|$ are the absolute value of the differences in the ex ante estimated beta on risk factor XYZ for the pair of firms. These are included in cross-sectional regressions with other explanatory variables found to be predictive of $\Delta \dot{\rho}_{ij}$. The standard errors, reported in parentheses below the regression coefficients, use the reshuffling methodology described in section 1.3. Statistical significance of the coefficient relative to the null hypothesis of zero is denoted using asterisks, where * indicates significance at the 5% level and ** indicates significance at the 1% level.

	(1)	(2)	(3)	(4)
1970's Decade Dummy	9.04* (3.98)	8.58* (3.97)	7.36 (4.48)	8.38* (3.90)
1980's Decade Dummy	2.27 (3.42)	2.25 (3.30)	2.18 (3.40)	2.97 (3.55)
1990's Decade Dummy	7.70** (2.94)	7.58* (2.95)	6.70* (3.11)	6.26* (3.07)
2000's Decade Dummy	5.31 (3.65)	5.85 (3.73)	5.77 (4.00)	6.43 (3.97)
Shared Trading Behavior (v_{ij})		-0.16 (0.10)	-0.22 (0.13)	-0.18 (0.13)
Same Exchange				-1.21* (0.59)
Shared Fundamentals ($\text{Corr}[ROE_i, ROE_j]$)			0.91** (0.22)	0.55* (0.22)
Same Sector				1.63** (0.44)
Same Industry				0.02 (0.63)
Same Subindustry				1.29 (0.90)
$ \beta_{i,MKT} - \beta_{j,MKT} $	-0.06 (0.34)	-0.03 (0.35)	-0.43 (0.51)	-0.36 (0.52)
$ \beta_{i,SMB} - \beta_{j,SMB} $	-0.68** (0.26)	-0.80** (0.26)	-0.38 (0.33)	-0.56 (0.33)
$ \beta_{i,HML} - \beta_{j,HML} $	-0.63* (0.26)	-0.64* (0.28)	-0.58 (0.34)	-0.43 (0.35)
Observations	6,148,574	4,397,326	2,207,508	1,947,768

microstructure effects, $\Delta \dot{p}$.

The regression results are summarized in Table 1.8. The first specification, with the difference in the betas on risk factors as the only explanatory variables shows the regression coefficients are negative—the opposite of our prediction. The coefficient for the difference in β_{MKT} is effectively zero.

In the other three regression specifications considered, the explanatory variables previously found to be significant are also included. The coefficients on the new variables measuring differences in risk factor loadings remain negative and hardly precise enough to distinguish from zero. It appears that time variation in discount rates in loadings on known risk factors may explain a small portion of the differences in long-horizon versus short-horizon correlations across this sample of US stocks, but this is not the sort of mean-reverting behavior commonly modeled and it is primarily driven by SMB and HML, not the aggregate equity market.

It should also be noted that the regression coefficients on the differences in risk exposures are certainly underestimated because of estimation error. This attenuation bias similarly affects the shared trading behavior and ROE correlation variables, which likely have even more estimation error than the betas on the risk factors.

1.5 IMPLICATIONS FOR ASSET PRICES AND INVESTORS

There is nothing about the proposed framework analyzing correlation and time horizon that is specific to the returns of individual stocks. In a traditional asset pricing context, we can consider how the time horizon will affect betas on risk factors, and hence, asset pricing.

As a first pass, consider how the return horizon affects the volatilities and correlations of the three factors of the Fama-French model. These are plotted in Figure 1.3 using the same time period as in the other empirical analysis, 1970-2009. Since these factor returns coexist for a much longer history than the typical equity security, we can consider long-term horizons that extend much longer than 6 months.

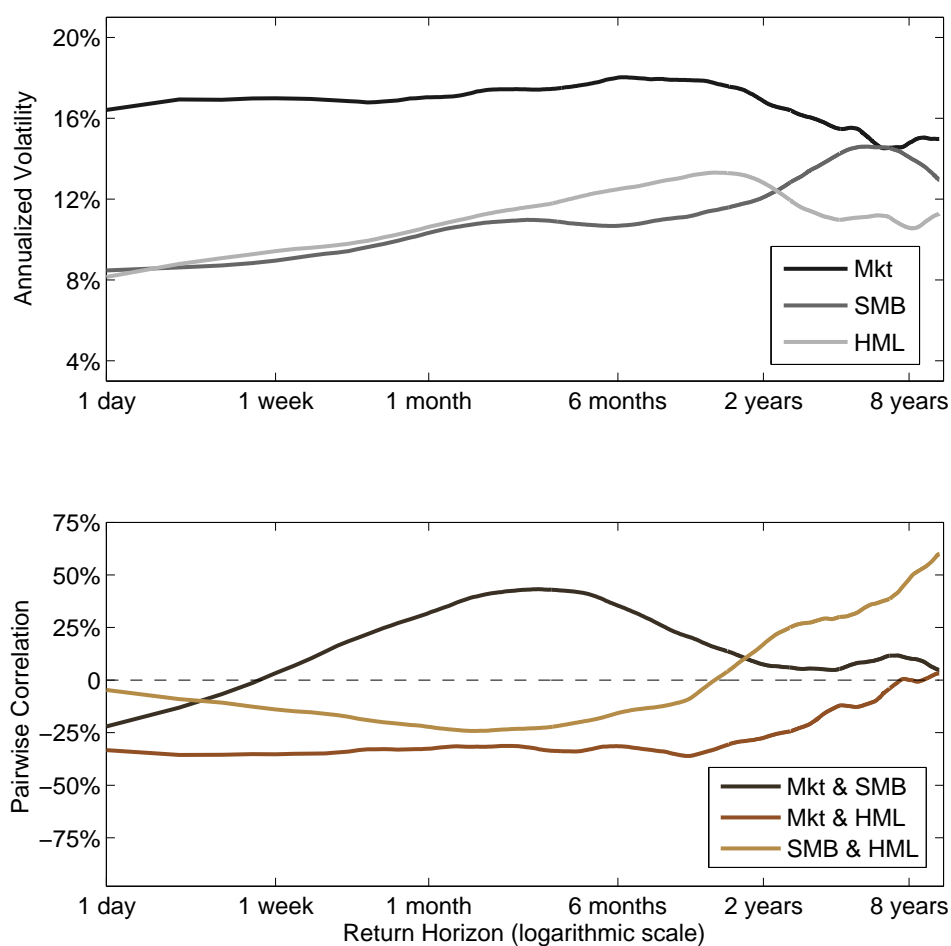


Figure 1.3: Annualized Volatility and Correlations for Risk Factors, 1970-2009

Looking at the top axis, plotting the estimate of volatility as a function of time horizon, the most striking feature is the upward sloping relationship for SMB and HML. The positive relationship between volatility and time horizon suggest that returns to the SMB and HML portfolios exhibit positive autocorrelation—at least at horizons in the range of 0-2 years. This is exactly the sort of behavior that would lead to the negative regression coefficients in the regression presented in Table 1.8. At the two year horizon, the HML volatility begins to decrease while the volatility of the SMB portfolio continues to increase for return horizons as long 6 or 7 years. This is indicative of momentum, rather than mean reversion, over these horizons.

Consistent with previous research (Fama and French, 1988), the broad market portfolio shows relatively little predictability for horizons shorter than one year, with a relatively constant relationship between volatility and time horizon. This would explain why aggregate market exposure explains little of the cross-sectional differences in $\Delta \hat{\rho}_{ij}$ at the stock level. The well-documented tendency for the aggregate stock market to exhibit mean reversion over long horizons begins to kick in as the horizon increases beyond one year.

The pairwise correlations are plotted on the lower axis in Figure 1.3. The SMB and HML portfolios have a negative relationship with the market portfolio over short and medium horizons, but these correlations tend toward zero as the return horizon lengthens. Perhaps the most striking relationship is the correlation between SMB and HML. While these portfolios seem to have uncorrelated returns over short horizons, the correlation coefficient increases significantly over long horizons. Repeating the caveat that estimates of long-horizon correlations can be noisy, the initial evidence

suggests that SMB and HML may be distinct risks over short time horizons but contain similar fundamental risks that become evident over longer time periods.

At the same time, the SMB and HML portfolios are not nearly as attractive to a long-horizon investor. While at horizons of a few days these portfolios seem to have half the volatility as the market portfolio, the volatility almost doubles when the horizon stretches to a few years. Worse still, these portfolios that previously seemed to offer good diversification relative to the aggregate equity market see their correlations increase significantly.

IMPLICATIONS FOR SHORT-TERM TRADERS

While buy-and-hold investors may have poor measures of risk calculated from short-horizon returns, active investors with a short-term focus (or even long-term investors who rebalance frequently) may find short-term comovement estimates appropriately capture the portfolio risks that matter to them. Although the underlying driver of short-horizon comovement may be fads rather than fundamentals, it accurately reflects the one-period risks they face.

However, the relationship between correlation and time horizon reveals how one period affects the next. As equation (1.8) emphasizes, correlation differences imply predictability. With predictability, there is an implied trading strategy that should be attractive to tactical traders.

In this section, I will show the historical performance of a simple trading strategy based on the comovement patterns identified. This exercise provides additional evidence that the comovement patterns established in the empirical analysis cannot

be easily explained by established risk factors. It also frames the results in a setting familiar to other empirical studies of asset (mis)pricing where a portfolio formation rule generates a trading strategy.

For better or for worse, this trading strategy based on comovement patterns has no anchor suggesting the true fundamental value of any particular asset. The intuition is roughly equivalent to that of a "pairs trading" strategy (albeit with a much longer horizon). When the prices of two assets with similar fundamentals diverge, the strategy puts on a long-short convergence trade. This comes with some danger. A more savvy investor would consider the actual news and prices rather than pursue what Stein (2009) terms an "unanchored" trading scheme. In that sense, the trading strategy is empirically instructive but not recommended.

A SIMPLE TRADING SIGNAL

The proposed trading signal is derived from the regression relationships for the short run return

$$E[r_{t,i}|r_{t,j}] = E[r_{t,i}] + \rho_{ij} (1) \frac{\sigma_i}{\sigma_j} (r_{t,j} - E[r_{t,j}]) \quad (1.21)$$

and the long run return

$$E\left[\sum_{\tau=0}^{H-1} r_{i,t+\tau} | r_{t,j}\right] = E\left[\sum_{\tau=0}^{H-1} r_{i,t+\tau}\right] + \rho_{ij} (H) \frac{\sigma_i}{\sigma_j} (r_{t,j} - E[r_{t,j}]) \quad (1.22)$$

of $r_{t,i}$ conditional on $r_{t,j}$. If we assume that the volatility ratio $(\frac{\sigma_i}{\sigma_j})$ is roughly constant and the unconditional expected return for each stock is approximately equal, then we

can subtract (1.21) from (1.22) and forecast the excess return for the future

$$E \left[\sum_{\tau=1}^{H-1} r_{i,t+\tau} | r_{t,j} \right] - E \left[\sum_{\tau=1}^{H-1} r_{i,t+\tau} \right] = \Delta \rho_{ij} \frac{\sigma_i}{\sigma_j} r_{t,j}. \quad (1.23)$$

With N assets, equation (1.23) will yield $N - 1$ univariate forecasts. For simplicity, the trading signal will weight them equally.³ The signal is then defined as

$$X_{i,t} = \frac{1}{N-1} \sum_{j \neq i} \Delta \rho_{ij} \frac{\sigma_i}{\sigma_j} r_{t,j}. \quad (1.24)$$

EMPIRICAL IMPLEMENTATION

In the empirical implementation of the trading strategy, the universe of firms will be determined in much the same way as before, comprising the 2000 largest firms by market cap over the 40 year sample. The set of firms will be updated annually, using data available the final business day in December of the previous year.

To predict the future difference in long-run and short-run correlation ($\Delta \rho_{i,j}$) I use the two main variables presented previously, where investor trading behavior is proxied by the correlation in bid-ask bounce, ν_{ij} , and fundamentals are measured as the correlation of the return on equity, $\text{Corr}[ROE_i, ROE_j]$. The difference between long-horizon and short-horizon correlation that determines the trading signal for forecasting in (1.24) can be constructed without too much fear of overfitting from the in-sample regression results by simply taking the equal-weighted difference:

$$\Delta \rho_{i,j} \approx \text{Corr}[ROE_i, ROE_j] - \nu_{ij}.$$

³An alternative would be to create the multivariate optimal forecast with GLS weights

These two variables are updated annually and implemented in portfolios formed each January using information that would be available in December. The volatility ratio $\frac{\sigma_i}{\sigma_j}$ is also updated annually, and is calculated as the standard deviation of the weekly returns over the prior three years. Shorter histories are used for any firm where three years are not available, and outliers are winsorized at the 5th and 95th percentiles.

SIGNAL PERSISTENCE

There remains the question of how long this signal should persist. The empirical analysis arbitrarily chose the long horizon to be $H = 26$ weeks but did not suggest whether the correlation differences resolved in a matter of weeks or if the correlations continued to evolve even after the six-month window. In the context of this trading strategy, this question is analogous to asking how long the signal X_t is expected to forecast excess returns.

In the framework of the simple model of fads and fundamentals presented earlier, we want to know the decay rates δ_d and δ_f . While there is likely a high degree of variation in the characteristics of fads and fundamentals that affect the US equity market, it is interesting to take the simplified model and estimate the half-life of the signal.

We can do this by building a simple portfolio rule, sorting stocks into quintiles based on their signal X_t and constructing a long-short portfolio that buys the highest quintile and sells the lowest quintile. The event time returns to this portfolio, shown in Figure 1.4 will show the degree to which the information persists.

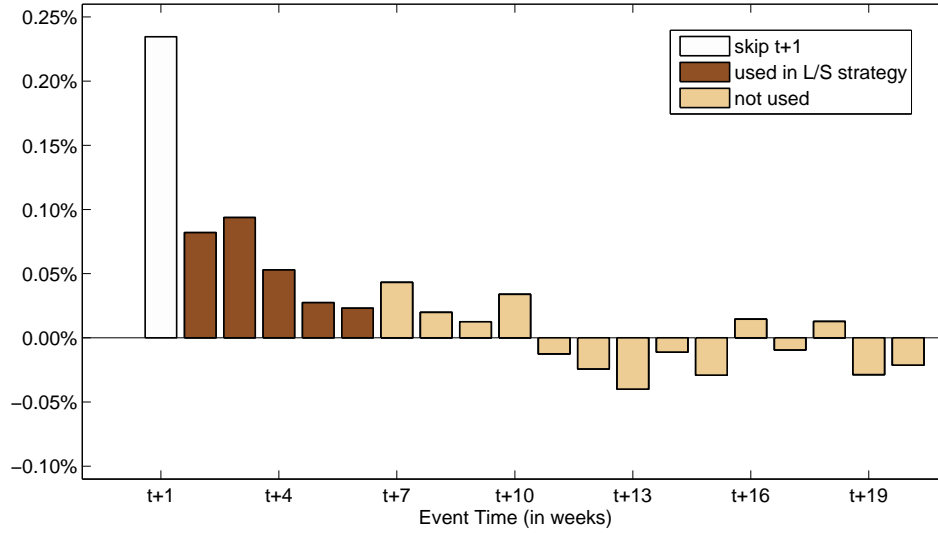


Figure 1.4: Event Time Returns to X_t Components of the L/S Portfolio

Each column in the bar chart represents the average weekly return resulting from a portfolio formed at time $t = 0$. The first column, colored white, represents the return that would be received from buying at the close of the formation week. To be as conservative as possible in representing the returns to a trading strategy, this first week is omitted from the trading strategy results shown in the following subsection. Even discarding this first week, there is a pattern of positive returns that continues at lags of up to two months.

BACKTEST RESULTS

Given the matrix $\Delta\beta$, the trading signal in (1.24) is obtained each week by multiplying $\Delta\beta$ by the returns from the recent past. For the purposes of this backtest, I will consider the recent past to be the returns from the past 6 weeks, omitting the

most recent weeks' returns to avoid the gaining credit for returns previously shown to be partially attributed to microstructure effects. The results without lagging the signal by one week would be extraordinarily large.

I generate calendar time backtest returns by sorting stocks each week into equal-weighted quintile portfolios based on their respective trading signal predicting future returns. The 200 firms with the highest factor values, populating portfolio Q_5 , are predicted to outperform the quintiles with lower factor values, particularly those in the quintile with the lowest factor values, Q_1 . A long/short portfolio is created by taking a long position in the firms in Q_5 and an equivalent short position in the firms comprising Q_1 .

I will also show event time returns that would result from creating the trading signals using only one week returns over a range of lags. This will give an indication of how fast the predicted components of excess and insufficient comovement are corrected in asset prices. This will also confirm the choice of using a six week window in the calendar time backtest is both sensible and robust to alternative specifications.

TRADING STRATEGY RESULTS

The annual returns to the long/short portfolio are graphed in Figure 1.5. The performance of this long/short portfolio is relatively consistent over time and does not show a tendency to decrease over time. This is true even in the most recent decade when you might expect that trading by hedge funds, especially so-called statistical arbitrage funds, might employ similar strategies and erode the returns available to a comovement based strategy.

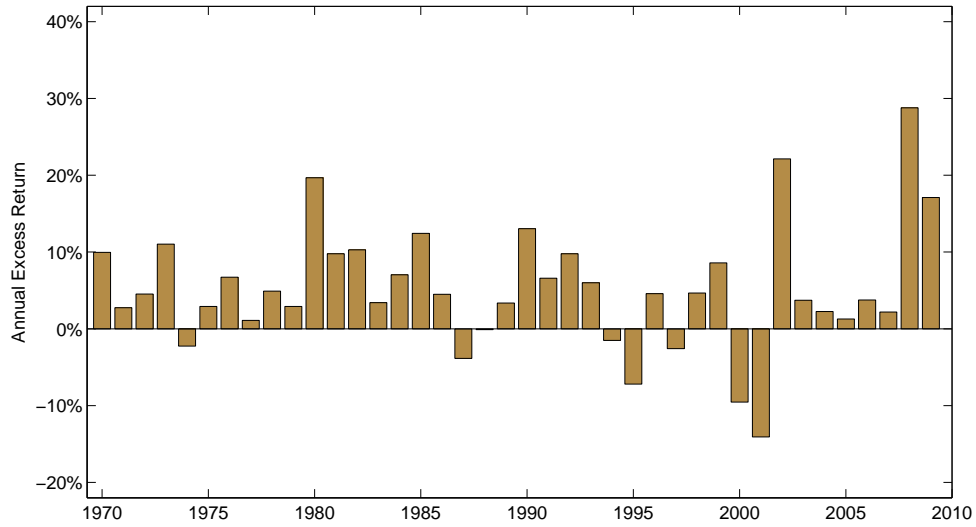


Figure 1.5: Annual Calendar Time Returns to L/S Portfolio

The strong recent performance is also surprising given the fact that, on average, short-horizon and long-horizon comovement have converged. This result suggests that the dispersion of comovement differences across firms remains large and predictable even while the average is near zero. Looking again at the annual returns to the strategy, the most profitable of the 40 years considered was 2008, with a return of 28.8%. Over the 40-year sample, the long/short portfolio generates an average annual excess return of 5.3% with a corresponding Sharpe Ratio of 0.65.

The weekly event time returns, shown in Figure 1.4, provide additional insight on the nature of the portfolio returns. These event time returns only interact one week of past returns (dated r_t) to generate the signal vector X_t . The event time graph displays the mean return to the long/short portfolio traded various weeks into the future. You can see that the $t + 1$ return is shaded in white. This is because the week immediately

following portfolio formation is excluded in the analysis, since some of that (very large) return may be generated by temporary price impact and would not be achievable. The returns from $t + 2$ to $t + 6$ are shaded in dark blue. This is to indicate that these five weeks of returns are the ones used in the construction of the calendar time long/short portfolios. Returns to all subsequent weeks are in light blue. From these event returns, it appears that the predictive component of comovement identified by these two signals generates declining abnormal returns for about 10 weeks after portfolio formation, and afterwards, the returns seem indistinguishable from noise.

ADJUSTING THE CALENDAR TIME RETURNS FOR RISK EXPOSURES

The average weekly excess returns alphas for the calendar time analysis of the five quintile portfolios and the long/short portfolio are presented in Table 1.9. As would be desired, there is a consistent pattern of returns increasing by quintile. In the unadjusted excess returns, the lowest quintile portfolio earns only 0.46 basis points per week versus the 8.6 basis point average return of the highest quintile, which corresponds to an annual return of 0.24%. The 8 basis point weekly return of the long/short portfolio has an associated t-statistic of 3.36, indicating we can confidently reject the notion that the true excess return of the strategy is zero.

Table 1.9 also reports the alphas for each portfolio after controlling for risk factors known to generate positive returns. These alphas are the intercept in the regression of the weekly returns of risk factors on the returns to the quintile and long/short portfolios. Four factor models are considered, and the Tuesday-to-Tuesday weekly returns for each of the component risk factors are derived from the daily research

Table 1.9: Weekly Abnormal Returns (in bps) to $\Delta\beta$ Trading Strategy

This table shows the calendar time portfolio abnormal returns, reported in basis points ($1/100^{\text{th}}$ of one percent). The first row shows the average weekly returns of the quintile portfolios and the long/short (L/S) portfolio formed by going long the highest quintile with the highest signal values (Q_5) and short the quintile portfolio with the lowest. Alpha is the intercept coefficient from regressing the weekly returns on various risk factors. The return series of the risk factors and the risk free rates are derived from the data provided by Ken French on his website. T-statistics are displayed in brackets below each return coefficient.

	Factor Quintile					L/S $Q_5 - Q_1$
	(low) Q_1	Q_2	Q_3	Q_4	(high) Q_5	
Excess Returns	0.46 [0.26]	2.24 [1.63]	4.36 [3.25]	6.54 [4.86]	8.62 [5.40]	8.16 [3.36]
1-factor alpha (Mkt)	-0.48 [-0.19]	4.45 [2.43]	7.78 [4.39]	9.72 [5.31]	9.95 [4.20]	10.43 [4.31]
3-factor alpha (...+ SMB, HML)	-4.14 [-2.10]	0.07 [0.05]	2.88 [2.12]	4.71 [3.42]	5.65 [3.36]	9.79 [4.02]
4-factor alpha (...+ UMD)	0.46 [0.26]	2.24 [1.63]	4.36 [3.25]	6.54 [4.86]	8.62 [5.40]	8.16 [3.36]
6-factor alpha (...+ STREV, LTREV)	-6.37 [-3.54]	-0.28 [-0.19]	2.12 [1.51]	5.92 [4.18]	9.03 [5.37]	15.40 [6.17]

returns available on Ken French's website⁴. The first two models include a 1-factor model that controls for exposure to the value-weighted market index, and the 3-factor alpha, that additionally includes the SMB and HML factors popularized by Fama and French (1992).

In addition to these standard benchmarks, we might wonder if the returns to portfolios based on comovement are related to momentum and reversal patterns found to empirically generate positive returns in the cross-section of US equities. To answer this, we can introduce two additional models, a 4-factor model including Carhart's (1997) momentum factor, and finally, a 6-factor model which additionally includes short-term and long-term reversal patterns. These reversal returns are defined by French to be the lagged one month return and the past 5-year return excluding the most recent year. Interestingly, this comovement trading strategy tends to trade in the opposite direction of these reversal factors, making the alphas look even more compelling. The long/short portfolio, which averages 8.2 basis points of excess returns weekly, reports a 6-factor alpha of 15.4 basis points. Translated to an annual time frequency, these risk adjusted returns would yield an average return of 8.4% and a Sharpe Ratio of 1.03.

CONCLUSION

Asset price comovement changes with time horizon. The evidence is consistent with a model where fads and information delays cause prices to temporarily deviate from fundamentals. In particular, there is compelling evidence that investor trading

⁴http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

behavior and salient security characteristics are more important factors in determining the correlation of US equity returns over short horizons while measures of long-run fundamentals play a greater role in return correlations over longer horizons.

I propose the difference between short-horizon and long-horizon comovement is a natural metric for studying excess comovement. Measures of common trading behavior and shared economic fundamentals show significant power in explaining cross-sectional differences in excess comovement across pairs of stocks. They can also form a successful trading strategy. A portfolio based on predictable differences in stock correlations generates consistent excess returns not explained by risk exposures.

The main implication for investors with a buy and hold strategy is that they may be underestimating (or overestimating) the risk concentration of their portfolio if they extrapolate comovement and volatility from short-horizon returns. This also suggests a degree of caution to financial econometricians who propose the use of intra-day data to estimate the covariance of security returns. It would seem that using ever shorter return horizons to estimate second moments will likely capture a greater degree of comovement driven by trading behavior rather than the fundamentals that matter over longer horizons.

Although the empirical evidence presented here focuses on US equities, the principle should apply just as much in other asset classes as well as in the broader asset allocation decision. In fact, there is reason to believe differences in comovement may be even larger across asset classes, as market segmentation may be more pronounced. The relationship between correlation and return horizon may identify

risks and opportunities that can arise as short-run comovement deviates from long-run fundamentals.

We must include in the long-period cost a third term which we might call the risk-cost to cover the unknown possibilities of the actual yield differing from the expected yield.

John Maynard Keynes

2

An Intertemporal CAPM with Stochastic Volatility

authored with John Campbell, Stefano Giglio and Christopher Polk

THE FUNDAMENTAL INSIGHT of intertemporal asset pricing theory is that long-term investors should care just as much about the returns they earn on their invested wealth as about the level of that wealth. In a simple model with a constant rate of return, for example, the sustainable level of consumption is the return on wealth multiplied by the level of wealth, and both terms in this product are equally

important. In a more realistic model with time-varying investment opportunities, conservative long-term investors will seek to hold “intertemporal hedges”, assets that perform well when investment opportunities deteriorate. Such assets should deliver lower average returns in equilibrium if they are priced from conservative long-term investors’ first-order conditions.

Since the seminal work of Merton (1973) on the intertemporal capital asset pricing model (ICAPM), a large empirical literature has explored the relevance of intertemporal considerations for the pricing of financial assets in general, and the cross-sectional pricing of stocks in particular. One strand of this literature uses the approximate accounting identity of Campbell and Shiller (1988a) and the assumption that a representative investor has Epstein-Zin utility (Epstein and Zin 1989) to obtain approximate closed-form solutions for the ICAPM’s risk prices (Campbell 1993). These solutions can be implemented empirically if they are combined with vector autoregressive (VAR) estimates of asset return dynamics (Campbell 1996). Campbell and Vuolteenaho (2004), Campbell, Polk, and Vuolteenaho (2010), and Campbell, Giglio, and Polk (2011) use this approach to argue that value stocks outperform growth stocks on average because growth stocks do well when the expected return on the aggregate stock market declines; in other words, growth stocks have low risk premia because they are intertemporal hedges for long-term investors.

A weakness of the papers cited above is that they ignore time-variation in the volatility of stock returns. In general, investment opportunities may deteriorate either because expected stock returns decline or because the volatility of stock returns

increases, and it is an empirical question which of these two types of intertemporal risk have a greater effect on asset returns. We address this weakness in this paper by extending the approximate closed-form ICAPM to allow for stochastic volatility. The resulting model explains risk premia in the stock market using three priced risk factors corresponding to three important attributes of aggregate market returns: revisions in expected future cash flows, discount rates, and volatility. An attractive characteristic of the model is that the prices of these three risk factors depend on only one free parameter, the long-horizon investor's coefficient of risk aversion.

Since the long-horizon investor in our model cares mostly about persistent changes in the investment opportunity set, there must be predictable variation in long-run volatility for volatility risk to matter. Empirically, we implement our methodology using a vector autoregression (VAR) including stock returns, realized variance, and other financial indicators that may be relevant for predicting returns and risk. Our VAR reveals low-frequency movements in market volatility tied to the default spread, the yield spread of low-rated over high-rated bonds. While this effect has received little attention in the literature, we argue that it is sensible: Investors in risky bonds perceive the long-run component of volatility and incorporate this information when they set credit spreads, as risky bonds are short the option to default. Moreover, we show that GARCH-based methods that filter only the information in past returns in order to disentangle the short-run and long-run volatility components miss this important low-frequency component.

With our novel model of long-run volatility in hand, we find that growth stocks have low average returns because they outperform not only when the expected stock

return declines, but also when stock market volatility increases. Thus growth stocks hedge two types of deterioration in investment opportunities, not just one. In the period since 1963 that creates the greatest empirical difficulties for the standard CAPM, we find that the three-beta model explains over 69% of the cross-sectional variation in average returns of 25 portfolios sorted by size and book-to-market ratios. The model is not rejected at the 5% level while the CAPM is strongly rejected. The implied coefficient of relative risk aversion is an economically reasonable 9.63, in contrast to the much larger estimate of 20.70, which we get when we estimate a comparable version of the two-beta CAPM of Campbell and Vuolteenaho (2004) using the same data.¹ This success is due in large part to the inclusion of volatility betas in the specification. In particular, the spread in volatility betas in the cross-section generates an annualized spread in average returns of 6.52% compared to a comparable spread of 3.90% and 2.24% for cash-flow and discount-rate betas.

We confirm that our findings are robust by expanding the set of test portfolios in two important dimensions. First, we show that our three-beta model not only describes the cross-section of size- and book-to-market-sorted portfolios but also can explain the average returns on risk-sorted portfolios. We examine risk-sorted portfolios in response to the argument of Daniel and Titman (1997, 2012) and Lewellen, Nagel, and Shanken (2010) that asset-pricing tests using only portfolios sorted by characteristics known to be related to average returns, such as size and value, can be misleading. As tests that include risk-sorted portfolios are unable to reject our intertemporal CAPM with stochastic volatility, we verify that the model's

¹The risk aversion estimate reported in Campbell and Vuolteenaho's (2004) paper is 28.75.

success is not simply due to the low-dimensional factor structure of the 25 size- and book-to-market-sorted portfolios. Specifically, we show that sorts on stocks' pre-formation sensitivity to volatility news generate economically and statistically significant spread in both post-formation volatility beta and average returns in a manner consistent with our model. Interestingly, in the post-1963 period, sorts on past CAPM beta generate little spread in post-formation cash-flow betas, but significant spread in post-formation volatility betas. Since, in the three-beta model, covariation with aggregate volatility news has a negative premium, the three-beta model also explains why stocks with high past CAPM betas have offered relatively little extra return in the post-1963 sample.

Second, we show that our three-beta model can help explain average returns on non-equity portfolios that are exposed to aggregate volatility risk. These portfolios include the S&P 100 index straddle of Coval and Shumway (2001), which is explicitly designed to be highly correlated with aggregate volatility risk, and the risky bond factor of Fama and French (1993), which should be sensitive to changes in aggregate volatility since risky corporate debt is short the option to default. Consistent with this intuition, we find that compared to the volatility beta of a value-minus-growth bet, the risky bond factor's volatility beta is of the same order of magnitude while the straddle's volatility beta is more than 3 times larger in absolute magnitude. These volatility betas are of the right sign to explain the abnormal CAPM returns of the option and bond portfolios. Approximately 38% of the average straddle return can be attributed to its three ICAPM betas, based purely on model estimates from the cross-section of equity returns. Additionally, when we price the *joint* cross-section of

equity, bond, and straddle returns our intertemporal CAPM with stochastic volatility is not rejected at the 5-percent level while the CAPM is strongly rejected.

Our work is complementary to recent research on the “long-run risk model” of asset prices (Bansal and Yaron 2004) which can be traced back to insights in Kandel and Stambaugh (1991). Both the approximate closed-form ICAPM and the long-run risk model start with the first-order conditions of an infinitely lived Epstein-Zin representative investor. As originally stated by Epstein and Zin (1989), these first-order conditions involve both aggregate consumption growth and the return on the market portfolio of aggregate wealth. Campbell (1993) pointed out that the intertemporal budget constraint could be used to substitute out consumption growth, turning the model into a Merton-style ICAPM. Restoy and Weil (1998, 2011) used the same logic to substitute out the market portfolio return, turning the model into a generalized consumption CAPM in the style of Breeden (1979).

Kandel and Stambaugh (1991) were the first researchers to study the implications for asset returns of time-varying first and second moments of consumption growth in a model with a representative Epstein-Zin investor. Specifically, Kandel and Stambaugh (1991) assumed a four-state Markov chain for the expected growth rate and conditional volatility of consumption, and provided closed-form solutions for important asset-pricing moments. In the spirit of Kandel and Stambaugh (1991), Bansal and Yaron (2004) added stochastic volatility to the Restoy-Weil model, and subsequent research on the long-run risk model has increasingly emphasized the importance of stochastic volatility for generating empirically plausible implications from this model (Bansal, Kiku, and Yaron 2012, Beeler and Campbell 2012). In this

paper we give the approximate closed-form ICAPM the same capability to handle stochastic volatility that its cousin, the long-run risk model, already possesses.

One might ask whether there is any reason to work with an ICAPM rather than a consumption-based model given that these models are derived from the same set of assumptions. The ICAPM developed in this paper has several advantages. First, it describes risks as they appear to an investor who takes asset prices as given and chooses consumption to satisfy his budget constraint. This is the way risks appear to individual agents in the economy, and it seems important for economists to understand risks in the same way that market participants do rather than relying exclusively on a macroeconomic perspective. Second, the ICAPM allows an empirical analysis based on financial proxies for the aggregate market portfolio rather than on accurate measurement of aggregate consumption. While there are certainly challenges to the accurate measurement of financial wealth, financial time series are generally available on a more timely basis and over longer sample periods than consumption series. Third, the ICAPM in this paper is flexible enough to allow multiple state variables that can be estimated in a VAR system; it does not require low-dimensional calibration of the sort used in the long-run risk literature. Finally, the stochastic volatility process used here governs the volatility of all state variables, including itself. We show that this assumption fits financial data reasonably well, and it guarantees that stochastic volatility would always remain positive in a continuous-time version of the model, a property that does not hold in most current implementations of the long-run risk model.²

²Eraker (2008) and Eraker and Shaliastovich (2008) are exceptions.

The closest precursors to our work are unpublished papers by Chen (2003) and Sohn (2010). Both papers explore the effects of stochastic volatility on asset prices in an ICAPM setting but make strong assumptions about the covariance structure of various news terms when deriving their pricing equations. Chen (2003) assumes constant covariances between shocks to the market return (and powers of those shocks) and news about future expected market return variance. Sohn (2010) makes two strong assumptions about asset returns and consumption growth, specifically that all assets have zero covariance with news about future consumption growth volatility and that the conditional contemporaneous correlation between the market return and consumption growth is constant through time. Duffee (2005) presents evidence against the latter assumption. It is in any case unattractive to make assumptions about consumption growth in an ICAPM that does not require accurate measurement of consumption.

Chen estimates a VAR with a GARCH model to allow for time variation in the volatility of return shocks, restricting market volatility to depend only on its past realizations and not those of the other state variables. His empirical analysis has little success in explaining the cross-section of stock returns. Sohn uses a similar but more sophisticated GARCH model for market volatility and tests how well short-run and long-run risk components from the GARCH estimation can explain the returns of various stock portfolios, comparing the results to factors previously shown to be empirically successful. In contrast, our paper incorporates the volatility process directly in the ICAPM, allowing heteroskedasticity to affect and to be predicted by all state variables, and showing how the price of volatility risk is pinned down by the

time-series structure of the model along with the investor's coefficient of risk aversion.

A working paper by Bansal, Kiku, Shaliastovich and Yaron (2012), contemporaneous with our own, explores the effects of stochastic volatility in the long-run risk model. Like us, they find stochastic volatility to be an important feature in the time series of equity returns. Their work puts greater emphasis on the implied consumption dynamics while we focus on the cross-sectional pricing implications of exposure to volatility news. More fundamentally, there are differences in the underlying models. They assume that the stochastic process driving volatility is homoskedastic, and in their cross-sectional analysis they impose that changes in the equity risk premium are driven only by the conditional variance of the stock market. The different modeling assumptions account for our contrasting empirical results; we show that volatility risk is very important in explaining the cross-section of stock returns while they find it has little impact on cross-sectional differences in risk premia.

Stochastic volatility has, of course, been explored in other branches of the finance literature. For example, Chacko and Viceira (2005) and Liu (2007) show how stochastic volatility affects the optimal portfolio choice of long-term investors. Chacko and Viceira assume an $AR(1)$ process for volatility and argue that movements in volatility are not persistent enough to generate large intertemporal hedging demands. Campbell and Hentschel (1992), Calvet and Fisher (2007), and Eraker and Wang (2011) argue that volatility shocks will lower aggregate stock prices by increasing expected returns, if they do not affect cash flows. The strength of this volatility feedback effect depends on the persistence of the volatility process. Coval

and Shumway (2001), Ang, Hodrick, Xing, and Zhang (2006), and Adrian and Rosenberg (2008) present evidence that shocks to market volatility are priced risk factors in the cross-section of stock returns, but they do not develop any theory to explain the risk prices for these factors.

There is also an enormous literature in financial econometrics on modeling and forecasting time-varying volatility. Since Engle's (1982) seminal paper on ARCH, much of the literature has focused on variants of the univariate GARCH model (Bollerslev 1986), in which return volatility is modeled as a function of past shocks to returns and of its own lags (see Poon and Granger (2003) and Andersen et al. (2006) for recent surveys). More recently, realized volatility from high-frequency data has been used to estimate stochastic volatility processes (Barndorff-Nielsen and Shephard 2002, Andersen et al. 2003). The use of realized volatility has improved the modeling and forecasting of volatility, including its long-run component; however, this literature has primarily focused on the information content of high-frequency intra-daily return data. This allows very precise measurement of volatility, but at the same time, given data availability constraints, limits the potential to use long time series to learn about long-run movements in volatility. In our paper, we measure realized volatility only with daily data, but augment this information with other financial time series that reveal information investors have about underlying volatility components.

A much smaller literature has, like us, looked directly at the information in other variables concerning future volatility. In early work, Schwert (1989) links movements in stock market volatility to various indicators of economic activity, particularly the

price-earnings ratio and the default spread, finding relatively weak results. Engle, Ghysels and Sohn (2009) study the effect of inflation and industrial production growth on volatility, finding a significant link between the two, especially at long horizons. Campbell and Taksler (2003) look at the cross-sectional link between corporate bond yields and equity volatility, emphasizing that bond yields respond to idiosyncratic firm-level volatility as well as aggregate volatility. Two recent papers, Paye (2012) and Christiansen et al. (2012), look at larger sets of potential predictors of volatility, that include the default spread and/or valuation ratios, to study which ones have predictive power for quarterly realized variance. The former, in a standard regression framework, finds that a few variables, that include the commercial paper to Treasury spread and the default spread, contain useful information for predicting volatility. The latter uses Bayesian Model Averaging to determine which variables are most important for predicting quarterly volatility, and documents the importance of the default spread and valuation ratios in forecasting short-run volatility.

2.1 AN INTERTEMPORAL MODEL WITH STOCHASTIC VOLATILITY

ASSET PRICING WITH TIME VARYING RISK

Preferences

We begin by assuming a representative agent with Epstein-Zin preferences. We write the value function as

$$V_t = \left[(1 - \delta) C_t^{\frac{1-\gamma}{\theta}} + \delta (E_t [V_{t+1}^{1-\gamma}])^{1/\theta} \right]^{\frac{\theta}{1-\gamma}}, \quad (2.1)$$

where C_t is consumption and the preference parameters are the discount factor δ , risk aversion γ , and the elasticity of intertemporal substitution ψ . For convenience, we define $\theta = (1 - \gamma)/(1 - 1/\psi)$.

The corresponding stochastic discount factor (SDF) can be written as

$$M_{t+1} = \left(\delta \left(\frac{C_t}{C_{t+1}} \right)^{1/\psi} \right)^\theta \left(\frac{W_t - C_t}{W_{t+1}} \right)^{1-\theta}, \quad (2.2)$$

where W_t is the market value of the consumption stream owned by the agent, including current consumption C_t .³ The log return on wealth is $r_{t+1} = \ln(W_{t+1}/(W_t - C_t))$, the log value of wealth tomorrow divided by reinvested wealth today. The log SDF is therefore

$$m_{t+1} = \theta \ln \delta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1) r_{t+1}. \quad (2.3)$$

A convenient identity

The gross return to wealth can be written

$$1 + R_{t+1} = \frac{W_{t+1}}{W_t - C_t} = \left(\frac{C_t}{W_t - C_t} \right) \left(\frac{C_{t+1}}{C_t} \right) \left(\frac{W_{t+1}}{C_{t+1}} \right), \quad (2.4)$$

expressing it as the product of the current consumption payout, the growth in consumption, and the future price of a unit of consumption.

We find it convenient to work in logs. We define the log value of reinvested wealth per unit of consumption as $z_t = \ln((W_t - C_t)/C_t)$, and the future value of a

³This notational convention is not consistent in the literature. Some authors exclude current consumption from the definition of current wealth.

consumption claim as $h_{t+1} = \ln (W_{t+1}/C_{t+1})$, so that the log return is:

$$r_{t+1} = -z_t + \Delta c_{t+1} + h_{t+1}. \quad (2.5)$$

Heuristically, the return on wealth is negatively related to the current value of reinvested wealth and positively related to consumption growth and the future value of wealth. The last term in equation (2.5) will capture the effects of intertemporal hedging on asset prices, hence the choice of the notation h_{t+1} for this term.

The ICAPM

We assume that asset returns are jointly conditionally lognormal, but we allow changing conditional volatility so we are careful to write second moments with time subscripts to indicate that they can vary over time. Under this standard assumption, the expected return on any asset must satisfy

$$0 = \ln E_t \exp\{m_{t+1} + r_{i,t+1}\} = E_t [m_{t+1} + r_{i,t+1}] + \frac{1}{2} \text{Var}_t [m_{t+1} + r_{i,t+1}], \quad (2.6)$$

and the risk premium on any asset is given by

$$E_t r_{i,t+1} - r_{f,t} + \frac{1}{2} \text{Var}_t r_{t+1} = -\text{Cov}_t [m_{t+1}, r_{i,t+1}]. \quad (2.7)$$

The convenient identity (2.5) can be used to write the log SDF (2.3) without reference to consumption growth:

$$m_{t+1} = \theta \ln \delta - \frac{\theta}{\psi} z_t + \frac{\theta}{\psi} h_{t+1} - \gamma r_{t+1}. \quad (2.8)$$

Since the first two terms in (2.5) are known at time t , only the latter two terms appear in the conditional covariance in (2.7). We obtain an ICAPM pricing equation that relates the risk premium on any asset to the asset's covariance with the wealth return and with shocks to future consumption claim values:

$$E_t r_{i,t+1} - r_{f,t} + \frac{1}{2} \text{Var}_t r_{t+1} = \gamma \text{Cov}_t [r_{i,t+1}, r_{t+1}] - \frac{\theta}{\psi} \text{Cov}_t [r_{i,t+1}, h_{t+1}] \quad (2.9)$$

Return and risk shocks in the ICAPM

To better understand the intertemporal hedging component h_{t+1} , we proceed in two steps. First, we approximate the relationship of h_{t+1} and z_{t+1} by taking a loglinear approximation about \bar{z} :

$$h_{t+1} \approx \kappa + \rho z_{t+1} \quad (2.10)$$

where the loglinearization parameter $\rho = \exp(\bar{z}) / (1 + \exp(\bar{z})) \approx 1 - C/W$.

Second, we apply the general pricing equation (2.6) to the wealth portfolio itself (setting $r_{i,t+1} = r_{t+1}$), and use the convenient identity (2.5) to substitute out consumption growth from this expression. Rearranging, we can write the variable z_t as

$$z_t = \psi \ln \delta + (\psi - 1) E_t r_{t+1} + E_t h_{t+1} + \frac{\psi}{\theta} \frac{1}{2} \text{Var}_t [m_{t+1} + r_{t+1}]. \quad (2.11)$$

Third, we combine these expressions to obtain the innovation in h_{t+1} :

$$\begin{aligned} h_{t+1} - E_t h_{t+1} &= \rho(z_{t+1} - E_t z_{t+1}) \\ &= (E_{t+1} - E_t) \rho \left(\begin{aligned} &(\psi - 1)r_{t+2} + h_{t+2} \\ &+ \frac{\psi}{\theta} \frac{1}{2} \text{Var}_{t+1} [m_{t+2} + r_{t+2}] \end{aligned} \right). \end{aligned} \quad (2.12)$$

Solving forward to an infinite horizon,

$$\begin{aligned} h_{t+1} - E_t h_{t+1} &= (\psi - 1)(E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\ &\quad + \frac{1}{2} \frac{\psi}{\theta} (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j \text{Var}_{t+j} [m_{t+1+j} + r_{t+1+j}] \\ &= (\psi - 1)N_{DR,t+1} + \frac{1}{2} \frac{\psi}{\theta} N_{RISK,t+1}. \end{aligned} \quad (2.13)$$

The second equality follows Campbell and Vuolteenaho (2004) and uses the notation N_{DR} (“news about discount rates”) for revisions in expected future returns. In a similar spirit we write revisions in expectations of future risk (the variance of the future log return plus the log stochastic discount factor) as N_{RISK} .

Finally, we substitute back into the intertemporal model (2.9):

$$\begin{aligned} E_t r_{i,t+1} - r_{f,t} + \frac{1}{2} \text{Var}_t r_{i,t+1} &= \gamma \text{Cov}_t [r_{i,t+1}, N_{CF,t+1}] \\ &\quad + \text{Cov}_t [r_{i,t+1}, -N_{DR,t+1}] \\ &\quad - \frac{1}{2} \text{Cov}_t [r_{i,t+1}, N_{RISK,t+1}]. \end{aligned} \quad (2.14)$$

This comes from the classic expression expressing the risk premium as risk

aversion γ times covariance with the current market return, plus $(\gamma - 1)$ times covariance with news about future market returns, minus one half covariance with risk. This is an extension of the ICAPM as written by Campbell (1993), with no reference to consumption or the elasticity of intertemporal substitution ψ .⁴ When the investor's risk aversion is greater than 1, assets which hedge aggregate discount rates ($\text{Cov}_t[r_{i,t+1}, N_{DR,t+1}] < 0$) or aggregate risk ($\text{Cov}_t[r_{i,t+1}, N_{RISK,t+1}] > 0$) have lower expected returns, all else equal.

In the rewritten form of equation (2.14), the expression follows Campbell and Vuolteenaho (2004), by breaking the market return into cash-flow news and discount-rate news. Cash-flow news N_{CF} is defined by $N_{CF} = r_{t+1} - E_t r_{t+1} + N_{DR}$. The price of risk for cash-flow news is γ times greater than the price of risk for discount-rate news, hence Campbell and Vuolteenaho call betas with cash-flow news “bad betas” and those with discount-rate news “good betas” since they have lower risk prices in equilibrium. The third term in (2.14) shows the risk premium associated with exposure to news about future risks and did not appear in Campbell and Vuolteenaho's model, which assumed homoskedasticity. Not surprisingly, the coefficient is negative, indicating that an asset providing positive returns when risk expectations increase will offer a lower return on average.

⁴Campbell (1993) briefly considers the heteroskedastic case, noting that when $\gamma = 1$, $\text{Var}_t[m_{t+1} + r_{t+1}]$ is a constant. This implies that N_{RISK} does not vary over time so the stochastic volatility term disappears. Campbell claims that the stochastic volatility term also disappears when $\psi = 1$, but this is incorrect. When limits are taken correctly, N_{RISK} does not depend on ψ (except indirectly through the loglinearization parameter, ρ).

FROM RISK TO VOLATILITY

The risk shocks defined in the previous subsection are shocks to the conditional volatility of returns plus the stochastic discount factor, that is, the conditional volatility of risk-neutralized returns. We now make additional assumptions on the data generating process for stock returns that allow us to estimate the news terms. These assumptions imply that the conditional volatility of risk-neutralized returns is proportional to the conditional volatility of returns themselves.

Suppose the economy is described by a first-order VAR

$$\mathbf{x}_{t+1} = \bar{\mathbf{x}} + \mathbf{\Gamma} (\mathbf{x}_t - \bar{\mathbf{x}}) + \sigma_t \mathbf{u}_{t+1}, \quad (2.15)$$

where \mathbf{x}_{t+1} is an $n \times 1$ vector of state variables that has r_{t+1} as its first element, σ_{t+1}^2 as its second element, and $n - 2$ other variables that help to predict the first and second moments of aggregate returns. $\bar{\mathbf{x}}$ and $\mathbf{\Gamma}$ are an $n \times 1$ vector and an $n \times n$ matrix of constant parameters, and \mathbf{u}_{t+1} is a vector of shocks to the state variables normalized so that its first element has unit variance. The key assumption here is that a scalar random variable, σ_t^2 , equal to the conditional variance of market returns, also governs time-variation in the variance of all shocks to this system. Both market returns and state variables, including volatility itself, have innovations whose variances move in proportion to one another.

Given this structure, news about discount rates can be written as

$$\begin{aligned}
N_{DR,t+1} &= (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\
&= \mathbf{e}'_1 \sum_{j=1}^{\infty} \rho^j \mathbf{\Gamma}^j \sigma_t \mathbf{u}_{t+1} \\
&= \mathbf{e}'_1 \rho \mathbf{\Gamma} (\mathbf{I} - \rho \mathbf{\Gamma})^{-1} \sigma_t \mathbf{u}_{t+1} \tag{2.16}
\end{aligned}$$

Furthermore, our log-linear model will make the log SDF, m_{t+1} , a linear function of the state variables. Since all shocks to the SDF are then proportional to σ_t ,

$\text{Var}_t [m_{t+1} + r_{t+1}] \propto \sigma_t^2$. As a result, the conditional variance,

$\text{Var}_t [(m_{t+1} + r_{t+1}) / \sigma_t] = \omega_t$, will be a constant that does not depend on the state variables. Without knowing the parameters of the utility function, we can write

$\text{Var}_t [m_{t+1} + r_{t+1}] = \omega \sigma_t^2$ so that the news about risk, N_{RISK} , is proportional to news about market return variance, N_V .

$$\begin{aligned}
N_{RISK,t+1} &= (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j \text{Var}_{t+j} [r_{t+1+j} + m_{t+1+j}] \\
&= (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j (\omega \sigma_{t+j}^2) \\
&= \omega \rho \mathbf{e}'_2 \sum_{j=0}^{\infty} \rho^j \mathbf{\Gamma}^j \sigma_t \mathbf{u}_{t+1} \\
&= \omega \rho \mathbf{e}'_2 (\mathbf{I} - \rho \mathbf{\Gamma})^{-1} \sigma_t \mathbf{u}_{t+1} = \omega N_{V,t+1}. \tag{2.17}
\end{aligned}$$

Substituting (2.17) into (2.14), we obtain an empirically-testable intertemporal

CAPM with stochastic volatility:

$$\begin{aligned}
E_t r_{i,t+1} - r_{f,t} + \frac{1}{2} \text{Var}_t r_{i,t+1} &= \gamma \text{Cov}_t [r_{i,t+1}, N_{CF,t+1}] \\
&\quad + \text{Cov}_t [r_{i,t+1}, -N_{DR,t+1}] \\
&\quad - \frac{1}{2} \omega \text{Cov}_t [r_{i,t+1}, N_{V,t+1}], \quad (2.18)
\end{aligned}$$

where covariances with news about three key attributes of the market portfolio (cash flows, discount rates, and volatility) describe the cross section of average returns.

The parameter ω is a nonlinear function of the coefficient of relative risk aversion γ , as well as the VAR parameters and the loglinearization coefficient ρ , but it does not depend on the elasticity of intertemporal substitution ψ except indirectly through the influence of ψ on ρ .

By definition

$$\begin{aligned}
\omega \sigma_t^2 &= \text{Var}_t [m_{t+1} + r_{t+1}] \\
&= \text{Var}_t \left[\frac{\theta}{\psi} h_{t+1} + (1 - \gamma) r_{t+1} \right] \\
&= \text{Var}_t \left[\frac{\theta}{\psi} \left((\psi - 1) N_{DR,t+1} + \frac{1}{2} \frac{\psi}{\theta} \omega N_{V,t+1} \right) + (1 - \gamma) r_{t+1} \right] \\
&= \text{Var}_t \left[(1 - \gamma) N_{DR,t+1} + \frac{1}{2} \omega N_{V,t+1} + (1 - \gamma) r_{t+1} \right] \\
&= \text{Var}_t \left[(1 - \gamma) N_{CF,t+1} + \frac{1}{2} \omega N_{V,t+1} \right].
\end{aligned}$$

Therefore ω solves:

$$\begin{aligned} \omega \sigma_t^2 = & (1 - \gamma)^2 \text{Var}_t [N_{CF,t+1}] \\ & + \omega(1 - \gamma) \text{Cov}_t [N_{CF,t+1}, N_{V,t+1}] + \omega^2 \frac{1}{4} \text{Var}_t [N_{V,t+1}] . \end{aligned} \quad (2.19)$$

We can see two main channels through which γ affects ω . First, a higher risk aversion—given the underlying volatilities of all shocks—implies a more volatile stochastic discount factor m , and therefore a higher RISK. This effect is proportional to $(1 - \gamma)^2$, so it increases rapidly with γ . Second, there is a feedback effect on RISK through future risk: ω appears on the right-hand side of the equation as well. Given that in our estimation we find $\text{Cov}_t [N_{CF,t+1}, N_{V,t+1}] < 0$, this second effect makes ω increase even faster with γ .⁵

This equation can also be written directly in terms of the VAR parameters. If we define x_{CF} and x_V as the error-to-news vectors such that

$$\frac{1}{\sigma_t} N_{CF,t+1} = x_{CF} u_{t+1} = (e'_1 + e'_1 \rho \Gamma (I - \rho \Gamma)^{-1}) u_{t+1} \quad (2.20)$$

$$\frac{1}{\sigma_t} N_{V,t+1} = x_V u_{t+1} = (e'_2 \rho (I - \rho \Gamma)^{-1}) u_{t+1} \quad (2.21)$$

and define the covariance matrix of the residuals (scaled to eliminate stochastic

⁵Bansal, Kiku, Shaliastovich and Yaron (2012) derive a similar expression. The equivalent expression for ω in their case reduces to $(1 - \gamma)^2$ as they impose that the volatility process is homoskedastic and the conditional equity premium is driven solely by the stochastic volatility.

volatility) as $\Sigma = \text{Var}[\mathbf{u}_{t+1}]$, then ω solves

$$0 = \omega^2 \frac{1}{4} x_V \Sigma x'_V - \omega (1 - (1 - \gamma) x_{CF} \Sigma x'_V) + (1 - \gamma)^2 x_{CF} \Sigma x'_{CF} \quad (2.22)$$

This quadratic equation for ω has two solutions. This result is an artifact of our linear approximation of the Euler Equation, and the appendix shows that one of the solutions can be disregarded. This false solution is easily identified by its implication that ω becomes infinite as volatility shocks become small. The correct solution is

$$\omega = \frac{1 - (1 - \gamma) x_{CF} \Sigma x'_V}{\frac{1}{2} x_V \Sigma x'_V} - \frac{\sqrt{(1 - (1 - \gamma) x_{CF} \Sigma x'_V)^2 - (1 - \gamma)^2 (x_V \Sigma x'_V) (x_{CF} \Sigma x'_{CF})}}{\frac{1}{2} x_V \Sigma x'_V} \quad (2.23)$$

There is an additional disadvantage to the quadratic expression arising from our loglinearization. In the case where risk aversion, volatility shocks and cash flow shocks are large enough, as measured by the product $(1 - \gamma)^2 (x_V \Sigma x'_V) (x_{CF} \Sigma x'_{CF})$, equation (2.22) may deliver a complex rather than a real value for ω . While the conditional variance $\text{Var}_t[m_{t+1} + r_{t+1}]$ from which we define ω will be both real and finite, the loglinear approximation may not allow for a real solution in an economically important region of the parameter space. Given our VAR estimates of the variance and covariance terms, we find equation (2.22) yields a real solution as γ ranges from zero to 6.93.

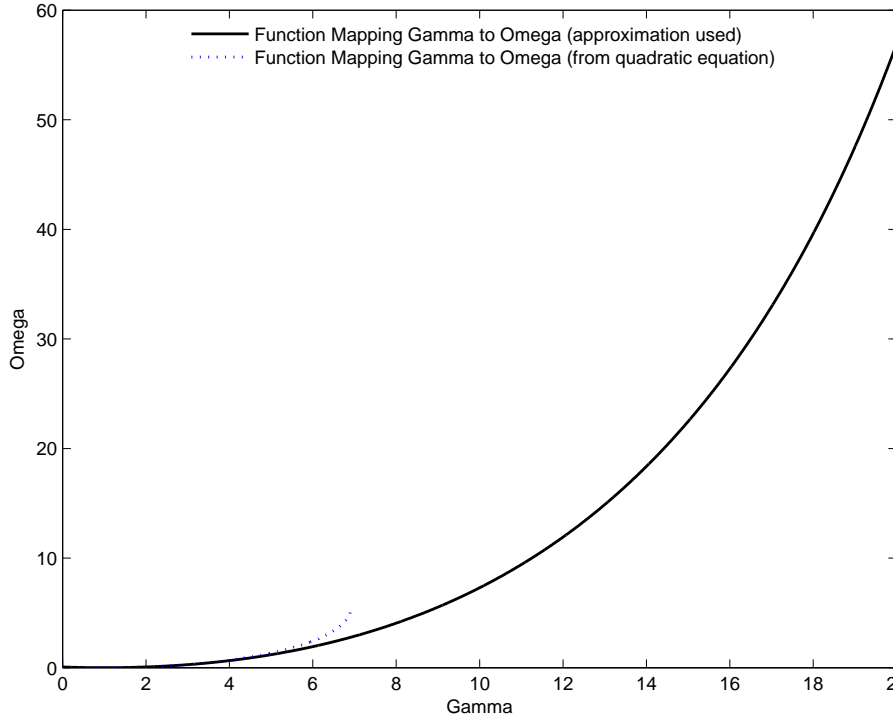


Figure 2.1: Approximate Gamma-Omega Relationship

This figure graphs the approximate relation between the parameter γ and the parameter ω described by equation (2.24) as well as the quadratic solution for ω described in equation (2.23). These functions depend on the loglinearization parameter ρ , set to 0.95 per year and the empirically estimated VAR parameters of Table 1. γ is the investor's risk aversion while ω is the sensitivity of news about risk, N_{RISK} , to news about market variance, N_V .

To allow for larger values in our risk aversion parameter, we consider an alternative approximation. If we linearize the right hand side of equation (2.19) around $\omega = 0$ we can approximate $\text{Var}_t[m_{t+1} + r_{t+1}]$ as a linear, rather than quadratic, function of ω . We then have

$$\omega \approx \frac{(1 - \gamma)^2 (x_{CF} \Sigma x'_{CF})}{1 - (1 - \gamma) (x_{CF} \Sigma x'_V)} \quad (2.24)$$

which is now defined for all $\gamma > 0$. Figure 2.1 plots ω as a function of γ using both the solution in equation (2.23) and the approximation in (2.24) for values of γ up to 20.

By construction, they will yield similar solutions for values of γ close to one, where ω gets close to 0 and volatility news becomes less and less important. In other words, it is easy to show that our linearization preserves the property of the true model that as $\gamma \rightarrow 1$, $\omega \rightarrow 0$ and

$$\text{Var}_t[m_{t+1} + r_{t+1}] \rightarrow (1 - \gamma)^2 \text{Var}_t[N_{CF}]$$

As risk aversion increases, we find that this approximate value for ω continues to resemble the exact solution of the quadratic equation (2.22) in the region where a real solution exists. We have also used numerical methods, similar to those proposed by Tauchen and Hussey (1991), to solve the model and validate our estimates of ω for a range of values for γ that include the region where the quadratic equation does not have a real solution.

IMPLICATIONS FOR CONSUMPTION GROWTH

Following Campbell (1993), in this paper we substitute consumption out of the pricing equations using the intertemporal budget constraint. However the model does have interesting implications for the implied consumption process. From equations (5) and (13), we can derive the expression:

$$\begin{aligned} \Delta c_{t+1} - E_t \Delta c_{t+1} &= (r_{t+1} - E_t r_{t+1}) - (\psi - 1) N_{DR,t+1} \\ &\quad - (\psi - 1) \frac{1}{2} \frac{\omega}{1 - \gamma} N_{V,t+1}. \end{aligned} \quad (2.25)$$

The first two components of the equation for consumption growth are the same as in the homoskedastic case. An unexpectedly high return of the wealth portfolio has a one-for-one effect on consumption. An increase in expected future returns increases today's consumption if $\psi < 1$, as the low elasticity of intertemporal substitution induces the representative investor to consume today (the income effect dominates). If $\psi > 1$, instead, the same increase induces the agent to reduce consumption to better exploit the improved investment opportunities (the substitution effect dominates).

The introduction of time-varying conditional volatility adds an additional term to the equation describing consumption growth. News about high future risk is news about a deterioration of future investment opportunities, which is bad news for a risk-averse investor ($\gamma > 1$). When $\psi < 1$, the representative agent will reduce consumption and save to ensure adequate future consumption. An investor with high elasticity of intertemporal substitution, on the other hand, will increase current consumption and reduce the amount of wealth exposed to the future (worse) investment opportunities.

Using estimates of the news terms from our VAR model (described in the next section), we can explore the implications of the model for consumption growth. As shown in the previous subsection, the three shocks that drive innovations in consumption growth ($r_{t+1} - E_t r_{t+1}$, $N_{DR,t+1}$, $N_{V,t+1}$) can all be expressed as functions of the vector of innovations $\sigma_t u_{t+1}$. The conditional variance of consumption growth, $\text{Var}_t(\Delta c_{t+1})$, will then be proportional to the conditional variance of returns, $\text{Var}_t(r_{t+1})$; similarly, the conditional standard deviation of consumption growth will be proportional to the conditional standard deviation of returns. As a consequence,

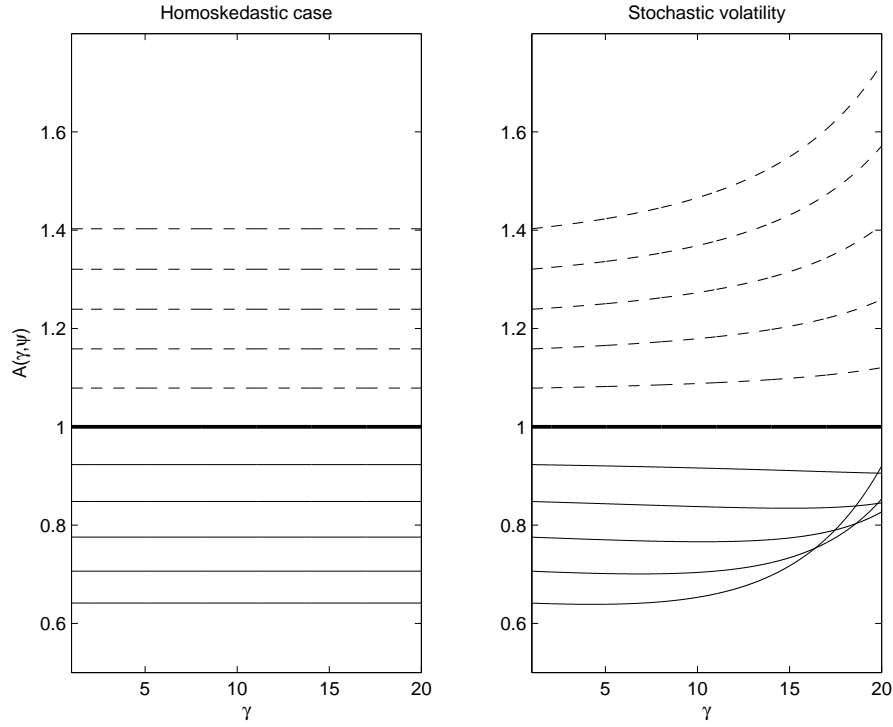


Figure 2.2: Consumption Growth Variance and Risk Aversion

This figure plots the coefficient $A(\gamma, \psi)$ relating the conditional volatility of consumption growth to the volatility of returns for different values of γ and ψ for the homoskedastic case (left panel) and for the heteroskedastic case (right panel), where $A(\gamma, \psi)$ is a function of the variances and covariances of the *scaled* residuals u_{t+1} . In each panel, we plot $A(\gamma, \psi)$ as γ varies between 1 and 20, for different values of ψ . Each line corresponds to a different ψ between 0.5 and 1.5.

the ratio of the standard deviations,

$$A(\gamma, \psi) \equiv \frac{\sqrt{\text{Var}_t(\Delta c_{t+1})}}{\sqrt{\text{Var}_t(r_{t+1})}}$$

will be a constant that depends on the model parameters γ and ψ as well as on the unconditional variances and covariances of the innovation vector u_{t+1} , which we obtain by estimating the VAR.

Figure 2.2 plots the coefficient $A(\gamma, \psi)$ for different values of γ and ψ for the homoskedastic case (left panel), and for the heteroskedastic case (right panel) using the linear approximation for ω described in Section 2.2. In each panel, we plot $A(\gamma, \psi)$ as γ varies between 0 and 20, for different values of ψ . Each line corresponds to a different ψ between 0.5 and 1.5; when $\psi = 1$ the value of $A(\gamma, \psi)$ is always equal to 1 since in that case the volatility of consumption growth is equal to the volatility of returns.

As expected, in the homoskedastic case (left panel), the variance of consumption growth does not depend on γ but only on ψ . It is rising in ψ because our VAR estimates imply that the return on wealth is negatively correlated with news about future expected returns $N_{DR,t+1}$, that is, wealth returns are mean-reverting. This confirms results reported in Campbell (1996). Once we add stochastic volatility (right panel), as γ increases the volatility of consumption growth increases for all values of ψ as long as $\psi \neq 1$. To understand why this is the case, notice in equation (2.24) that since ω grows with γ faster than $(1 - \gamma)^2$, the term $\frac{\omega}{1-\gamma}$ is increasing in γ in absolute value. Therefore, the larger γ , the more the variance of N_V gets amplified into a higher variance of consumption innovations.

Note also that for $\psi < 1$ and for high enough γ (i.e. in the bottom-right section of the right panel), the volatility of consumption innovations is *higher* for *lower* values of ψ . When risk aversion is high, innovations in consumption are dominated by news about future risk. Agents with very low or very high elasticity of intertemporal substitution, i.e. with ψ far from 1, will tend to adjust their consumption strongly (in different directions) to volatility news. Therefore, it is possible for individuals with

lower elasticity of intertemporal substitution to end up with a *more volatile* process for consumption innovations, due to their strong reaction to volatility news.

2.2 PREDICTING AGGREGATE STOCK RETURNS AND VOLATILITY

STATE VARIABLES

Our full VAR specification of the vector \mathbf{x}_{t+1} includes six state variables, five of which are the same as in Campbell, Giglio and Polk (2011). To those five variables, we add an estimate of conditional volatility. The data are all quarterly, from 1926:2 to 2011:4.

The first variable in the VAR is the log real return on the market, r_M , the difference between the log return on the Center for Research in Securities Prices (CRSP) value-weighted stock index and the log return on the Consumer Price Index.

The second variable is expected market variance ($EVAR$). This variable is meant to capture the volatility of market returns, σ_t , conditional on information available at time t , so that innovations to this variable can be mapped to the N_V term described above. To construct $EVAR_t$, we proceed as follows. We first construct a series of within-quarter *realized* variance of daily returns for each time t , $RVAR_t$. We then run a regression of $RVAR_{t+1}$ on lagged realized variance ($RVAR_t$) as well as the other five state variables at time t . This regression then generates a series of predicted values for $RVAR$ at each time $t + 1$, that depend on information available at time t : \widehat{RVAR}_{t+1} . Finally, we define our expected variance at time t to be exactly this predicted value at $t + 1$:

$$EVAR_t \equiv \widehat{RVAR}_{t+1}.$$

Note that though we describe our methodology in a two-step fashion where we first estimate *EVAR* and then use *EVAR* in a VAR, this is only for interpretability. Indeed, this approach to modeling *EVAR* can be considered a simple renormalization of equivalent results we would find from a VAR that included *RVAR* directly.⁶

The third variable is the price-earnings ratio (*PE*) from Shiller (2000), constructed as the price of the S&P 500 index divided by a ten-year trailing moving average of aggregate earnings of companies in the S&P 500 index. Following Graham and Dodd (1934), Campbell and Shiller (1988b, 1998) advocate averaging earnings over several years to avoid temporary spikes in the price-earnings ratio caused by cyclical declines in earnings. We avoid any interpolation of earnings as well as lag the moving average by one quarter in order to ensure that all components of the time- t price-earnings ratio are contemporaneously observable by time t . The ratio is log transformed.

Fourth, the term yield spread (*TY*) is obtained from Global Financial Data. We compute the *TY* series as the difference between the log yield on the 10-Year US Constant Maturity Bond (IGUSA10D) and the log yield on the 3-Month US Treasury Bill (ITUSA3D).

Fifth, the small-stock value spread (*VS*) is constructed from data on the six “elementary” equity portfolios also obtained from Professor French’s website. These elementary portfolios, which are constructed at the end of each June, are the intersections of two portfolios formed on size (market equity, *ME*) and three portfolios formed on the ratio of book equity to market equity (*BE/ME*). The size

⁶Since we weight observations based on *RVAR* in the first stage and then reweight observations using *EVAR* in the second stage, our two-stage approach in practice is not exactly the same as a one-stage approach. However, the results from a *RVAR*-weighted single-step estimation are qualitatively very similar to those produced by our two-stage approach.

breakpoint for year t is the median NYSE market equity at the end of June of year t . BE/ME for June of year t is the book equity for the last fiscal year end in $t - 1$ divided by ME for December of $t - 1$. The BE/ME breakpoints are the 30th and 70th NYSE percentiles.

At the end of June of year t , we construct the small-stock value spread as the difference between the $\ln(BE/ME)$ of the small high-book-to-market portfolio and the $\ln(BE/ME)$ of the small low-book-to-market portfolio, where BE and ME are measured at the end of December of year $t - 1$. For months from July to May, the small-stock value spread is constructed by adding the cumulative log return (from the previous June) on the small low-book-to-market portfolio to, and subtracting the cumulative log return on the small high-book-to-market portfolio from, the end-of-June small-stock value spread. The construction of this series follows Campbell and Vuolteenaho (2004) closely.

The sixth variable in our VAR is the default spread (DEF), defined as the difference between the log yield on Moody's BAA and AAA bonds. The series is obtained from the Federal Reserve Bank of St. Louis. Campbell, Giglio and Polk (2011) add the default spread to the Campbell and Vuolteenaho (2004) VAR specification in part because that variable is known to track time-series variation in expected real returns on the market portfolio (Fama and French, 1989), but mostly because shocks to the default spread should to some degree reflect news about aggregate default probabilities. Of course, news about aggregate default probabilities should in turn reflect news about the market's future cash flows.

SHORT-RUN VOLATILITY ESTIMATION

In order for the regression model that generates $EVAR_t$ to be consistent with a reasonable data-generating process for market variance, we deviate from standard OLS in two ways. First, we constrain the regression coefficients to produce fitted values (i.e. expected market return variance) that are positive. Second, given that we explicitly consider heteroskedasticity of the innovations to our variables, we estimate this regression using Weighted Least Squares (WLS), where the weight of each observation pair $(RVAR_{t+1}, \mathbf{x}_t)$ is initially based on the time- t value of $(RVAR)^{-1}$. However, to ensure that the ratio of weights across observations is not extreme, we shrink these initial weights towards equal weights. In particular, we set our shrinkage factor large enough so that the ratio of the largest observation weight to the smallest observation weight is always less than or equal to five. Though admittedly somewhat ad hoc, this bound is consistent with reasonable priors of the degree of variation over time in *expected* market return variance. More importantly, we show later that our results are robust to variation in this bound. Both the constraint on the regression's fitted values and the constraint on WLS observation weights bind in the sample we study.

The results of the first stage regression generating the state variable $EVAR_t$ are reported in Table 2.1 Panel A. Perhaps not surprisingly, past realized variance strongly predicts future realized variance. More importantly, the regression documents that an increase in either *PE* or *DEF* predicts higher future realized volatility. Both of these results are very statistically significant and are a novel finding of the paper. In

Table 2.1: VAR Estimation

The table shows the parameter estimates for a first-order VAR model. Panel A reports WLS estimates of a first-stage regression forecasting $RVAR$ with the state variables. Panel B reports WLS estimates of the full second-stage VAR. The first seven columns report coefficients on the explanatory variables, and the remaining column shows the R^2 and F statistics. Bootstrapped standard errors that take into account the uncertainty in generating $EVAR$ are in parentheses. The sample period for the dependent variables is 1926.3-2011.4, 342 quarterly data points.

Panel A: Forecasting Quarterly Realized Variance ($RVAR_{t+1}$)									
Second stage	Constant	$r_{M,t}$	$RVAR_t$	PE_t	TY_t	DEF_t	VS_t	$R^2\%/F$	
	-0.082 (0.033)	-0.016 (0.020)	0.394 (0.064)	0.023 (0.009)	-0.002 (0.002)	0.023 (0.006)	0.006 (0.010)	23.46% 18.42	
Panel B: VAR Estimates									
Second stage	Constant	$r_{M,t}$	$EVAR_t$	PE_t	TY_t	DEF_t	VS_t	$R^2\%/F$	
$r_{M,t+1}$	0.219 (0.120)	0.057 (0.068)	0.312 (0.571)	-0.054 (0.034)	0.004 (0.008)	-0.010 (0.022)	-0.032 (0.035)	3.32% 1.91	
$EVAR_{t+1}$	-0.065 (0.029)	-0.010 (0.005)	0.440 (0.097)	0.018 (0.008)	-0.001 (0.002)	0.016 (0.005)	0.007 (0.008)	47.28% 50.07	
PE_{t+1}	0.154 (0.116)	0.138 (0.064)	0.284 (0.546)	0.955 (0.033)	0.004 (0.007)	-0.011 (0.021)	-0.015 (0.033)	96.79% 1684.87	
TY_{t+1}	-0.047 (0.543)	-0.097 (0.336)	1.273 (2.789)	0.030 (0.157)	0.820 (0.035)	0.166 (0.111)	0.004 (0.160)	72.42% 146.63	
DEF_{t+1}	0.191 (0.263)	-0.383 (0.155)	1.649 (1.259)	-0.056 (0.074)	0.000 (0.017)	0.834 (0.051)	0.067 (0.077)	78.75% 206.96	
VS_{t+1}	0.138 (0.108)	0.075 (0.063)	0.762 (0.524)	-0.017 (0.031)	-0.004 (0.007)	-0.004 (0.021)	0.939 (0.031)	91.15% 575.20	

particular, the fact that we find that very persistent variables like *PE* and *DEF* forecast next period's volatility indicates a potential important role in volatility news for lower frequency or long-run movements in stochastic volatility.

We argue that the links we find are sensible. Investors in risky bonds incorporate their expectation of future volatility when they set credit spreads, as risky bonds are short the option to default. Therefore we expect higher *DEF* to be associated with higher *RVAR*. The result that higher *PE* predicts *higher RVAR* might seem surprising at first, but one has to remember that the coefficient indicates the effect of a change in *PE* holding constant the other variables, in particular the default spread. Since the default spread should also generally depend on the equity premium and since most of the variation in *PE* is due to variation in the equity premium, for a given value of the default spread, a relatively high value of *PE* implies a relatively higher level of future volatility. Thus *PE* cleans up the information in *DEF* concerning future volatility.

The R^2 of this regression is just over 23%. The relatively low R^2 masks the fact that the fit is indeed quite good, as we can see from Figure 2.3, in which *RVAR* and *EVAR* are plotted together. The R^2 is heavily influenced by the occasional spikes in realized variance, which the simple linear model we use is not able to capture. Indeed, our WLS approach downweights the importance of those spikes in the estimation procedure.

The internet appendix to this paper (Campbell, Giglio, Polk, and Turley 2012) reports descriptive statistics for these variables for the full sample, the early sample, and the modern sample. Consistent with Campbell, Giglio and Polk (2012), we document high correlation between *DEF* and both *PE* and *VS*. The table also

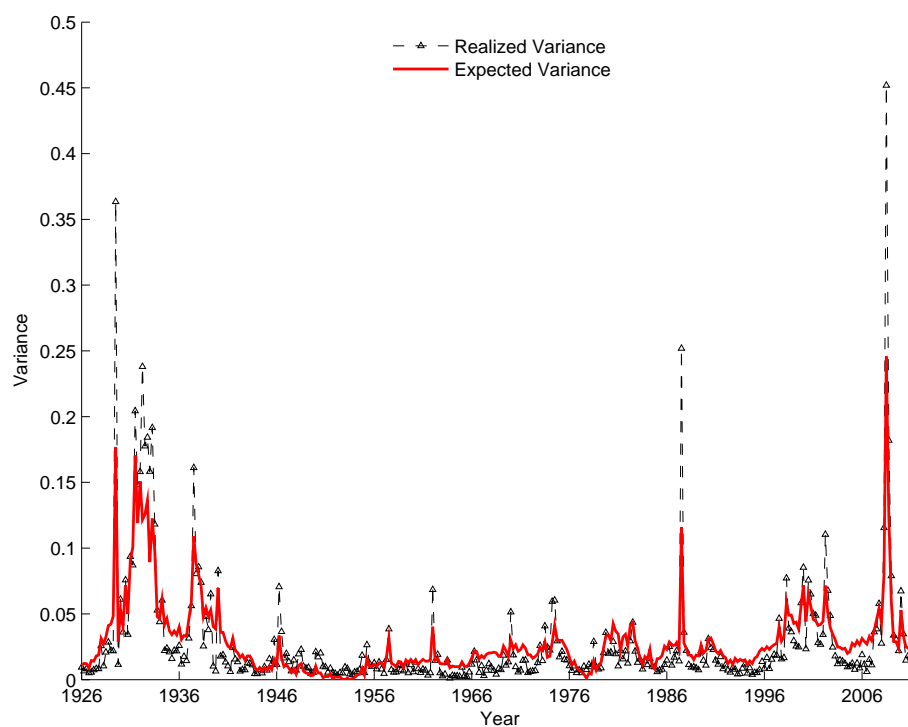


Figure 2.3: Realized and Expected Variance, 1926-2011

This figure plots quarterly observations of realized within-quarter daily return variance over the sample period 1926:2-2011:4 and the expected variance implied by the estimated model.

documents the persistence of both *RVAR* and *EVAR* (autocorrelations of 0.524 and 0.740 respectively) and the high correlation between these variance measures and the default spread.

Perhaps the most notable difference between the two subsamples is that the correlation between *PE* and several of our other state variables changes dramatically. In the early sample, *PE* is quite negatively correlated with both *RVAR* and *VS*. In the modern sample, *PE* is essentially uncorrelated with *RVAR* and quite positively correlated with *VS*. As a consequence, since *EVAR* is just a linear combination of our state variables, the correlation between *PE* and *EVAR* changes sign across the two samples. In the early sample, this correlation is very negative, with a value of -0.511. This strong negative correlation reflects the high volatility that occurred during the Great Depression when prices were relatively low. In the modern sample, the correlation is positive, 0.140. The positive correlation simply reflects the economic fact that episodes with high volatility and high stock prices, such as the technology boom of the late 1990s, were more prevalent in this subperiod than episodes with high volatility and low stock prices, such as the recession of the early 1980s.

ESTIMATION OF THE VAR AND THE NEWS TERMS

Following Campbell (1993), we estimate a first-order VAR as in equation (2.15), where \mathbf{x}_{t+1} is a 6×1 vector of state variables ordered as follows:

$$\mathbf{x}_{t+1} = [r_{M,t+1} \text{ } EVAR_{t+1} \text{ } PE_{t+1} \text{ } TY_{t+1} \text{ } DEF_{t+1} \text{ } VS_{t+1}]$$

so that the real market return $r_{M,t+1}$ is the first element and $EVAR$ is the second element. $\bar{\mathbf{x}}$ is a 6×1 vector of the means of the variables, and $\mathbf{\Gamma}$ is a 6×6 matrix of constant parameters. Finally, $\sigma_t \mathbf{u}_{t+1}$ is a 6×1 vector of innovations, with the conditional variance-covariance matrix of \mathbf{u}_{t+1} a constant:

$$\mathbf{\Sigma} = \text{Var}(\mathbf{u}_{t+1})$$

so that the parameter σ_t^2 scales the entire variance-covariance matrix of the vector of innovations.

The first-stage regression forecasting realized market return variance described in the previous section generates the variable $EVAR$. The theory in Section 2.2 assumes that σ_t^2 , proxied for by $EVAR$, scales the variance-covariance matrix of state variable shocks. Thus, as in the first stage, we estimate the second-stage VAR using WLS, where the weight of each observation pair $(\mathbf{x}_{t+1}, \mathbf{x}_t)$ is initially based on $(EVAR_t)^{-1}$. We continue to constrain both the weights across observations and the fitted values of the regression forecasting $EVAR$.

Table 2.1 Panel B presents the results of the VAR estimation for the full sample (1926:2 to 2011:4). We report bootstrap standard errors for the parameter estimates of the VAR that take into account the uncertainty generated by forecasting variance in the first stage. Consistent with previous research, we find that PE negatively predict future returns, though the t-statistic indicates only marginal significance. The value spread has a negative but not statistically significant effect on future returns. In our specification, a higher conditional variance, $EVAR$, is associated with higher future returns, though the effect is not statistically significant. Of course, the relatively high

degree of correlation among PE , DEF , VS , and $EVAR$ complicates the interpretation of the individual effect of those variables. As for the other novel aspects of the transition matrix, both high PE and high DEF predict higher future conditional variance of returns. High past market returns forecast lower $EVAR$, higher PE , and lower DEF .⁷

Tables 2.2 and 2.3 report the sample correlation and autocorrelation matrices of both the unscaled residuals $\sigma_t \mathbf{u}_{t+1}$ and the scaled residuals \mathbf{u}_{t+1} . The correlation matrices report standard deviations on the diagonals. There are a couple of aspects of these results to note. For one thing, a comparison of the standard deviations of the unscaled and scaled residuals provides a rough indication of the effectiveness of our empirical solution to the heteroskedasticity of the VAR. In general, the standard deviations of the scaled residuals are several times larger than their unscaled counterparts. More specifically, our approach implies that the scaled return residuals should have unit standard deviation. Our implementation results in a sample standard deviation of 0.562, that is relatively close to one.

Additionally, a comparison of the unscaled and scaled autocorrelation matrices reported in Table 2.3 reveals that much of the sample autocorrelation in the unscaled residuals is eliminated by our WLS approach. For example, the unscaled residuals in the regression forecasting the log real return have an autocorrelation of -0.074. The

⁷One worry is that many of the elements of the transition matrix are estimated imprecisely. Though these estimates may be zero, their non-zero but statistically insignificant in-sample point estimates, in conjunction with the highly-nonlinear function that generates discount-rate and volatility news, may result in misleading estimates of risk prices. However, the results are qualitatively similar if we instead employ a partial VAR where, via a standard iterative process, only variables with t -statistics greater than 1.0 are included in each VAR regression.

Table 2.2: VAR Residual Correlations and Standard Deviations

The table reports the correlation ("Corr/std") matrices of both the unscaled and scaled shocks from the second-stage VAR; the correlation matrix reports shock standard deviations on the diagonal. The sample period for the dependent variables is 1926.3-2011.4, 342 quarterly data points.

Corr/std	r_M	<i>EVAR</i>	<i>PE</i>	<i>TY</i>	<i>DEF</i>	<i>VS</i>
unscaled						
r_M	0.106	-0.488	0.907	-0.022	-0.489	-0.036
<i>EVAR</i>	-0.488	0.018	-0.575	-0.074	0.645	0.121
<i>PE</i>	0.907	-0.575	0.099	-0.011	-0.601	-0.064
<i>TY</i>	-0.022	-0.074	-0.011	0.561	0.006	-0.024
<i>DEF</i>	0.000	-0.489	0.645	-0.601	0.006	0.290
<i>VS</i>	-0.036	0.121	-0.064	-0.024	0.316	0.086
scaled						
r_M	0.568	-0.484	0.904	-0.043	-0.383	0.023
<i>EVAR</i>	-0.484	0.090	-0.561	-0.069	0.627	0.088
<i>PE</i>	0.904	-0.561	0.522	-0.033	-0.488	0.004
<i>TY</i>	-0.043	-0.069	-0.033	3.247	0.018	-0.033
<i>DEF</i>	-0.383	0.627	-0.488	0.018	1.363	0.261
<i>VS</i>	0.023	0.088	0.004	-0.033	0.261	0.496

Table 2.3: VAR Residual Autocorrelations

The table reports the autocorrelation ("Autocorr.") matrices of both the unscaled and scaled shocks from the second-stage VAR; the correlation matrix reports shock standard deviations on the diagonal. The sample period for the dependent variables is 1926.3-2011.4, 342 quarterly data points.

Autocorr.	$r_{M,t+1}$	$EVAR_{t+1}$	PE_{t+1}	TY_{t+1}	DEF_{t+1}	VS_{t+1}
unscaled						
$r_{M,t}$	-0.074	0.092	-0.067	0.047	0.100	0.045
$EVAR_t$	0.071	-0.153	0.083	-0.126	-0.183	-0.087
PE_t	-0.086	0.177	-0.151	0.070	0.221	0.093
TY_t	-0.046	0.075	-0.029	-0.088	0.081	0.050
DEF_t	0.152	-0.124	0.186	-0.157	-0.311	-0.147
VS_t	0.022	-0.034	0.020	-0.076	-0.080	-0.097
scaled						
$r_{M,t}$	0.002	0.045	-0.004	0.009	0.007	-0.006
$EVAR_t$	0.060	-0.102	0.073	-0.082	-0.120	-0.060
PE_t	-0.012	0.125	-0.077	0.027	0.109	0.027
TY_t	-0.036	0.067	-0.028	-0.058	0.073	0.039
DEF_t	0.094	-0.083	0.123	-0.111	-0.218	-0.107
VS_t	0.018	-0.031	0.009	-0.044	-0.066	-0.083

Table 2.4: VAR Specification Test

The table reports the results of regressions forecasting the squared second-stage residuals from the VAR estimated in Table 2.1 with $EVAR_t$. Bootstrap standard errors that take into account the uncertainty in generating $EVAR$ are in parentheses. The sample period for the dependent variables is 1926.3-2011.4, 342 quarterly data points.

Heteroskedastic Shocks			
Squared, second-stage, unscaled residual	Constant	$EVAR_t$	$R^2\%$
$r_{M,t+1}$	-0.003 (0.004)	0.478 (0.076)	19.78%
$EVAR_{t+1}$	0.000 (0.000)	0.018 (0.006)	5.86%
PE_{t+1}	-0.004 (0.004)	0.484 (0.076)	19.61%
TY_{t+1}	0.205 (0.084)	3.770 (1.837)	1.67%
DEF_{t+1}	-0.117 (0.044)	6.960 (0.922)	26.12%
VS_{t+1}	0.004 (0.002)	0.118 (0.034)	5.47%

corresponding autocorrelation of the scaled return residuals is essentially zero, 0.002.

Though the scaled residuals in the $EVAR$, PE and DEF regression still display some negative autocorrelation, the unscaled residuals are much more negatively autocorrelated.

Table 2.4 reports the coefficients of a regression of the squared unscaled residuals $\sigma_t u_{t+1}$ of each VAR equation on a constant and $EVAR$. These results are consistent with our assumption that $EVAR$ captures the conditional volatility of market returns (the coefficient on $EVAR$ in the regression forecasting the squared residuals of r_M is

0.478). The fact that *EVAR* significantly predicts with a positive sign all the squared errors of the VAR supports our underlying assumption that one parameter (σ_t^2) drives the volatility of all innovations.

The top panel of Table 2.5 presents the variance-covariance matrix and the standard deviation/correlation matrix of the news terms, estimated as described above. Consistent with previous research, we find that discount-rate news is twice as volatile as cash-flow news.

The interesting new results in this table concern the variance news term N_V . First, news about future variance is more volatile than discount-rate news. Second, it is negatively correlated (-0.22) with cash-flow news: as one might expect from the literature on the “leverage effect” (Black 1976, Christie 1982), news about low cash flows is associated with news about higher future volatility. Third, N_V correlates negatively (-0.09) with discount-rate news, indicating that news of high volatility tends to coincide with news of low future real returns.⁸ The net effect of these correlations, documented in the lower left panel of Table 2.5, is a slightly negative correlation of -.02 between our measure of volatility news and contemporaneous market returns (for related research see French, Schwert, and Stambaugh 1987).

The lower right panel of Table 2.5 reports the decomposition of the vector of innovations $\sigma_t^2 u_{t+1}$ into the three terms $N_{CF,t+1}$, $N_{DR,t+1}$, and $N_{V,t+1}$. As shocks to *EVAR* are just a linear combination of shocks to the underlying state variables, which includes *RVAR*, we “unpack” *EVAR* to express the news terms as a function of r_M , PE ,

⁸Though the point estimate is negative, the large standard errors imply that we cannot reject the “volatility feedback effect” (Campbell and Hentschel 1992, Calvet and Fisher 2007).

Table 2.5: Cash-flow, Discount-rate, and Variance News for the Market Portfolio

The table shows the properties of cash-flow news (N_{CF}), discount-rate news (N_{DR}), and volatility news (N_V) implied by the VAR model of Table 2.1. The upper sections show covariances on the left and the correlation matrix on the right, with standard deviations on the diagonal. The lower-left section shows the correlation of shocks to individual state variables with the news terms. The lower-right section shows the functions ($\mathbf{e}_1' + \mathbf{e}_1'\lambda_{DR}$, $\mathbf{e}_1'\lambda_{DR}$, $\mathbf{e}_2'\lambda_V$) that map the state-variable shocks to cash-flow, discount-rate, and variance news. Bootstrap standard errors that take into account the uncertainty in generating $EVAR$ are in parentheses.

News cov.	N_{CF}	N_{DR}	N_V	News corr/std	N_{CF}	N_{DR}	N_V
N_{CF}	0.00213 (0.00074)	-0.00042 (0.00106)	-0.00106 (0.00089)	N_{CF}	0.046 (0.007)	-0.101 (0.229)	-0.221 (0.239)
N_{DR}	-0.00042 (0.00106)	0.00823 (0.00261)	-0.00085 (0.00209)	N_{DR}	-0.101 (0.229)	0.091 (0.014)	-0.091 (0.350)
N_V	-0.00106 (0.00089)	-0.00085 (0.00209)	0.01074 (0.00312)	N_V	-0.221 (0.239)	-0.091 (0.350)	0.104 (0.021)
Shock correlations	N_{CF}	N_{DR}	N_V	Functions	N_{CF}	N_{DR}	N_V
r_M shock	0.523 (0.210)	-0.901 (0.036)	-0.019 (0.329)	r_M shock	0.924 (0.030)	-0.076 (0.030)	-0.051 (0.053)
$RVAR$ shock	-0.056 (0.143)	0.434 (0.106)	0.452 (0.150)	$RVAR$ shock	-0.092 (0.233)	-0.092 (0.233)	1.289 (0.414)
PE shock	0.180 (0.240)	-0.967 (0.035)	-0.090 (0.351)	PE shock	-0.856 (0.159)	-0.856 (0.159)	0.758 (0.282)
TY shock	0.104 (0.155)	0.078 (0.110)	-0.113 (0.227)	TY shock	0.010 (0.013)	0.010 (0.013)	-0.016 (0.023)
DEF shock	-0.160 (0.192)	0.490 (0.116)	0.741 (0.242)	DEF shock	-0.009 (0.036)	-0.009 (0.036)	0.314 (0.063)
VS shock	-0.435 (0.184)	-0.179 (0.138)	0.566 (0.262)	VS shock	-0.244 (0.125)	-0.244 (0.125)	0.412 (0.220)

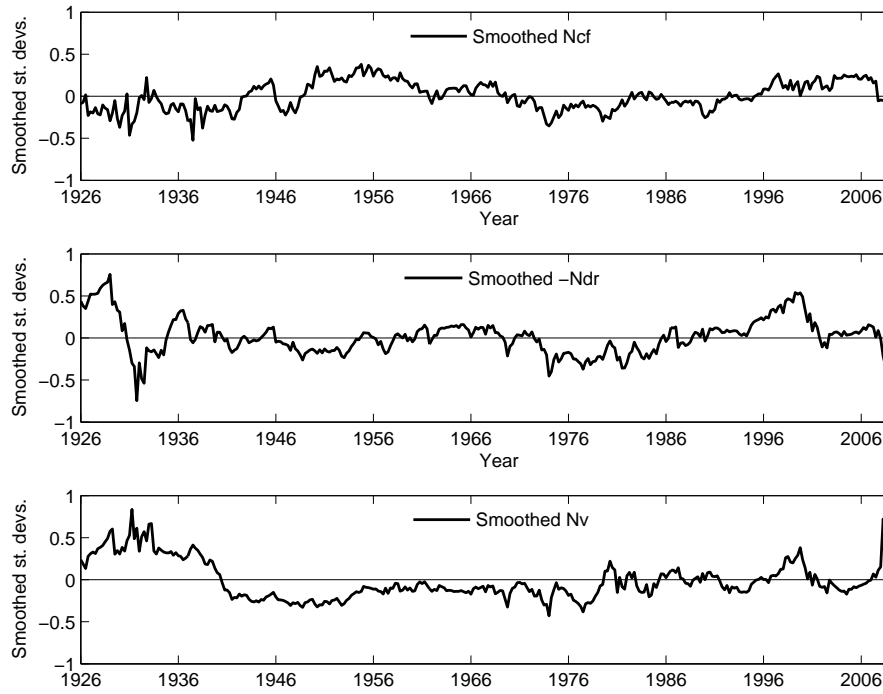


Figure 2.4: Normalized News Series

This figure plots normalized cash-flow news, the negative of normalized discount-rate news, and normalized variance news. The series are smoothed with a trailing exponentially-weighted moving average where the decay parameter is set to 0.08 per quarter, and the smoothed news series is generated as $MA_t(N) = 0.08N_t + (1 - 0.08)MA_{t-1}(N)$. This decay parameter implies a half-life of six years. The sample period is 1926:2-2011:4.

TY , VS , DEF , and $RVAR$. The panel shows that innovations to $RVAR$ are mapped more than one-to-one to news about future volatility. However, several of the other state variables also drive news about volatility. Specifically, we find that innovations in PE , DEF , and VS are associated with news of higher future volatility.

Figure 2.4 plots the smoothed series for N_{CF} , $-N_{DR}$ and N_V using an exponentially-weighted moving average with a quarterly decay parameter of 0.08. This decay parameter implies a half-life of six years. The pattern of N_{CF} and $-N_{DR}$ we

find is consistent with previous research. As a consequence, we focus on the smoothed series for market variance news. There is considerable time variation in N_V , and in particular we find episodes of news of high future volatility during the Great Depression and just before the beginning of World War II, followed by a period of little news until the late 1960s. From then on, periods of positive volatility news alternate with periods of negative volatility news in cycles of 3 to 5 years. Spikes in news about future volatility are found in the early 1970s (following the oil shocks), in the late 1970s and again following the 1987 crash of the stock market. The late 1990s are characterized by strongly negative news about future returns, and at the same time higher expected future volatility. The recession of the late 2000s is instead characterized by strongly negative cash-flow news, together with a spike in volatility of the highest magnitude in our sample. The recovery from the financial crisis has brought positive cash-flow news together with news about lower future volatility.

PREDICTING LONG-RUN VOLATILITY

The predictability of volatility, and especially of its long-run component, is central to this paper. In the previous sections, we have shown that volatility is strongly predictable, and it is predictable in particular by variables beyond lagged realizations of volatility itself: *PE* and *DEF* contain essential information about future volatility. We have also proposed a VAR-based methodology to construct long-horizon forecasts of volatility that incorporate all the information in lagged volatility as well as in the additional predictors like *PE* and *DEF*.

We now ask how well our proposed long-run volatility forecasts capture the

long-horizon component of volatility. In Table 2.6 we regress realized long-run variance up to period h ,

$$LHRVAR_h = \frac{\sum_{j=1}^h \rho^{j-1} RVAR_{t+j}}{\sum_{j=1}^h \rho^{j-1}},$$

on different forecasting models of long-run variance.⁹ In particular, we estimate two standard GARCH-type models, specifically designed to capture the long-run component of volatility. The first one is the two-component EGARCH model proposed by Adrian and Rosenberg (2008). This model assumes the existence of two separate components of volatility, one of which is more persistent than the other, and therefore will tend to capture the long-run dynamics of the volatility process. The other model we estimate is the FIGARCH model of Baillie, Bollerslev, and Mikkelsen (1996), in which the process for volatility is modeled as a fractionally-integrated process, and whose slow, hyperbolic rate of decay of lagged, squared innovations potentially captures long-run movements in volatility better. We first estimate both GARCH models using the full sample of daily returns and then generate the appropriate forecast of $LHRVAR_h$.¹⁰ To these two models, we add the set of variables from our VAR, and compare the forecasting ability of these different models.

Table 2.6 Panel A reports, for different horizons h ranging from 1 year to 15 years, the results of forecasting regressions of long run volatility $LHRVAR_h$ using different specifications. The first row of each sub-panel presents results using the state variables

⁹Note that we rescale by the sum of the weights ρ^j to maintain the scale of the coefficients in the predictive regressions across different horizons.

¹⁰We start our forecasting exercise in January 1930 so that we have a long enough history of past returns to feed the FIGARCH model.

in our VAR, each included separately. The second row predicts $LHRVAR_h$ with the horizon-specific forecast implied by our VAR (VAR_h). The third and fourth rows forecast $LHRVAR_h$ with the corresponding forecast from the EGARCH model (EG_h) and the FIGARCH model (FIG_h) respectively. The fifth and sixth rows join the VAR variables with the two GARCH-based forecasts, one at a time. The seventh and eighth row conducts a horse race between VAR_h and FIG_h and between VAR_h and DEF .

First note that both the EGARCH and FIGARCH forecasts by themselves capture a significant portion of the variation in long-run realized volatility: both have significant coefficients, and both have nontrivial R^2 s, even at very long horizons. Our VAR variables provide as good or better explanatory power, and $RVAR$, PE and DEF appear strongly statistically significant at all horizons (with the exception of $RVAR$ at $h = 20$, i.e. 5 years). Finally, the VAR-implied forecast, VAR_h , is not only significantly different from 0, but it is also not significantly different from 1. This indicates that our VAR is able to produce forecasts of volatility that not only go in the right direction, but are also of the right magnitude, even at very long horizons.

Very interesting results appear once we join our variables to the two GARCH models. Even after controlling for the GARCH-based forecasts (which render $RVAR$ insignificant), PE and DEF always come in significantly in predicting long-horizon volatility. Moreover, and especially at long horizons, the addition of the VAR state variables strongly increases the R^2 . We further show that when using the VAR-implied forecast together with the FIGARCH forecast, the coefficient on VAR_h is still very close to one and always statistically significant while the FIGARCH

Table 2.6: Forecasting Long-Horizon Realized Variance

The table reports the WLS parameter estimates of constrained regressions forecasting the annualized discounted sum of future $RVAR$ over the next 4 quarters ($\sum_{k=1}^h \rho^{(k-1)} RVAR_{t+k} / \sum_{k=1}^h \rho^{(k-1)}$). Initial WLS weights are inversely proportional to the corresponding FIG_h long-horizon forecast except in those regressions involving VAR_h or EG_h forecasts, where the corresponding VAR_h or EG_h long-horizon forecast is used instead. Newey-West standard errors estimated with lags corresponding to twice the number of overlapping observations are in square brackets. The sample period for the dependent variable is 1930.1-2011.4.

Panel A: Varying the Horizon h in $(\sum_{k=1}^h \rho^{(k-1)} RVAR_{t+k} / \sum_{k=1}^h \rho^{(k-1)})$										
Constant	r_M	$RVAR$	PE	TY	DEF	VS	VAR_h	EG_h	FIG_h	$R^2\%/F$
		$h = 4$ (1 years ahead)								
-0.089 [0.027]	-0.027 [0.021]	0.211 [0.094]	0.026 [0.008]	-0.002 [0.001]	0.025 [0.008]	0.007 [0.007]				31.24% 25.53
-0.002 [0.005]							0.998 [0.218]			31.33% 148.85
-0.007 [0.004]								1.054 [0.172]		34.38% 170.79
-0.001 [0.004]									0.998 [0.185]	25.01% 109.06
-0.071 [0.022]	-0.026 [0.018]	-0.132 [0.086]	0.018 [0.006]	-0.002 [0.001]	0.016 [0.006]	0.004 [0.007]		0.926 [0.220]		41.37% 33.66
-0.082 [0.026]	-0.025 [0.019]	-0.038 [0.092]	0.024 [0.008]	-0.002 [0.001]	0.020 [0.007]	0.003 [0.008]			0.679 [0.215]	34.41% 25.28
							0.713 [0.196]		0.507 [0.176]	34.16% 85.06
					0.005 [0.006]		0.857 [0.197]			28.86% 66.72

coefficient moves closer to zero (though estimates of the coefficient on FIG_h remain statistically significant at some horizons).

We develop an additional test of our VAR-based model of stochastic volatility from the idea that the variables that form the VAR – in particular the strongest of them, DEF – should predict volatility at long horizons only *through* the VAR, not *in addition* to it. In other words, the VAR forecasts should ideally represent the best way to combine the information contained in the state variables concerning long-run volatility. If true, after controlling for the VAR-implied forecast, DEF or other variables that enter the VAR should not significantly predict future long-run volatility. We test this hypothesis by running a regression using *both* the VAR-implied forecast and DEF as right-hand side variables. We find that at all horizons the coefficient on VAR_h is still not significantly different from 1, while the coefficient on DEF is small and statistically indistinguishable from 0.

Finally, in Panel B of Table 2.6 we examine more carefully the link between DEF and $LHRVAR$ focusing on the 10-year horizon. The Table reports the results from regressions forecasting $LHRVAR_{40}$ with PE , DEF , PEO (PE orthogonalized to DEF), and $DEFO$ (DEF orthogonalized to PE). The Table shows that by itself, PE has no information about low-frequency variation in volatility. In contrast, DEF forecasts nearly 22% of the variation in $LHRVAR_{40}$. And once DEF is orthogonalized to PE , the R^2 increases to 51%. Adding PEO has little effect on the R^2 . We argue that this is clear evidence of the strong predictive power of the orthogonalized component of the default spread.

Recall our simple interpretation of these results. DEF contains information about

future volatility as risky bonds are short the option to default. However, *DEF* also contains information about future aggregate risk premia. We know from previous work that most of the variation in *PE* is about aggregate risk premia. Therefore, including *PE* in the volatility forecasting regression cleans up variation in *DEF* due to aggregate risk premia and thus sharpens the link between *DEF* and future volatility. Since *PE* and *DEF* are negatively correlated (default spreads are relatively low when the market trades rich), both *PE* and *DEF* receive positive coefficients in the multiple regression.

In Figure 2.5, we provide a visual representation of the volatility-forecasting power of our key VAR state variables and our interpretation of the results. The top panel plots $LHRVAR_{40}$ together with lagged *DEF* and *PE*. The graph confirms the strong negative correlation between *PE* and *DEF* (correlation of -0.6) and highlights how both variables track long-run movements in long run volatility. To isolate the contribution of the default spread in predicting long run volatility, the bottom panel plots $LHRVAR_{40}$ together with *DEFO*. In general, the improvement in fit moving from the top panel to the bottom panel is clear.

More specifically, the contrasting behavior of *DEF* and *DEFO* in the two panels during episodes such as the tech boom help illustrate the workings of our story. Taken in isolation, the relatively stable default spread throughout most of the late 1990s would predict little change in expectations of future market volatility. However, once the declining equity premium over that period is taken into account (as shown by the rapid increase in *PE*), one recognizes that a *PE*-adjusted spread in the late 1990s actually forecasted much higher volatility ahead.

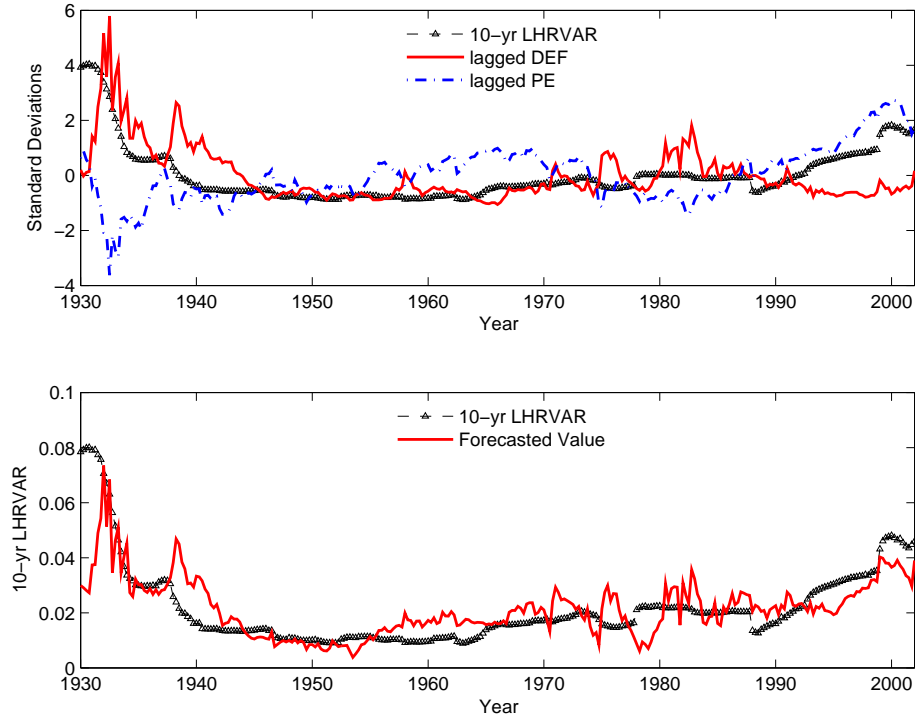


Figure 2.5: Key Components of Long-Horizon Volatility

We measure long-horizon realized variance ($LHRVAR$) as the annualized discounted sum of within-quarter daily return variance, $LHRVAR_h = \frac{\sum_{j=1}^h \rho^{j-1} RVAR_{t+j}}{\sum_{j=1}^h \rho^{j-1}}$. Each panel of this figure plots quarterly observations of ten-year realized variance, $LHRVAR_{40}$, over the sample period 1930:1-2001:1. In Panel A, in addition to $LHRVAR_{40}$, we also plot lagged PE and DEF . In Panel B, in addition to $LHRVAR_{40}$, we also plot the fitted value from a regression forecasting $LHRVAR_{40}$ with $DEFO$, defined as DEF orthogonalized to demeaned PE .

Taken together, the results in Table 2.1 Panel A and Table 2.6 make a strong case that credit spreads and valuation ratios contain information about future volatility not captured by simple univariate models, even those like the FIGARCH model or the two-component EGARCH model that are designed to fit long-run movements in volatility, and that our VAR method for calculating long-horizon forecasts preserves this information.

2.3 PRICING CASH-FLOW, DISCOUNT-RATE, AND VOLATILITY BETAS

TEST ASSETS

In addition to the six VAR state variables, our analysis also requires returns on a cross-section of test assets. We construct three sets of portfolios to use as test assets. Our primary cross-section consists of the excess returns on the 25 ME- and BE/ME-sorted portfolios, studied in Fama and French (1993), extended in Davis, Fama, and French (2000), and made available by Professor Kenneth French on his web site.¹¹

Daniel and Titman (1997, 2012) and Lewellen, Nagel, and Shanken (2010) point out that it can be misleading to test asset pricing models using only portfolios sorted by characteristics known to be related to average returns, such as size and value. In particular, characteristics-sorted portfolios are likely to show some spread in betas identified as risk by almost any asset pricing model, at least in sample. When the model is estimated, a high premium per unit of beta will fit the large variation in

¹¹<http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>

average returns. Thus, at least when premia are not constrained by theory, an asset pricing model may spuriously explain the average returns to characteristics-sorted portfolios.

To alleviate this concern, we follow the advice of Daniel and Titman (1997, 2012) and Lewellen, Nagel, and Shanken (2010) and construct a second set of six portfolios double-sorted on past risk loadings to market and variance risk. First, we run a loading-estimation regression for each stock in the CRSP database where $r_{i,t}$ is the log stock return on stock i for month t .

$$\sum_{j=1}^3 r_{i,t+j} = b_o + b_{r_M} \sum_{j=1}^3 r_{M,t+j} + b_{\Delta VAR} \sum_{j=1}^3 \Delta VAR_{t+j} + \varepsilon_{i,t+3}$$

We calculate ΔVAR as a weighted sum of changes in the VAR state variables. The weight on each change is the corresponding value in the linear combination of VAR shocks that defines news about market variance. We choose to work with changes rather than shocks as this allows us to generate pre-formation loading estimates at a frequency that is different from our VAR. Namely, though we estimate our VAR using calendar-quarter-end data, our approach allows a stock's loading estimates to be updated at each interim month.

The regression is reestimated from a rolling 36-month window of overlapping observations for each stock at the end of each month. Since these regressions are estimated from stock-level instead of portfolio-level data, we use quarterly data to minimize the impact of infrequent trading. With loading estimates in hand, each month we perform a two-dimensional sequential sort on market beta and ΔVAR

beta. First, we form three groups by sorting stocks on \hat{b}_{r_M} . Then, we further sort stocks in each group to three portfolios on $\hat{b}_{\Delta VAR}$ and record returns on these nine value-weight portfolios. The final set of risk-sorted portfolios are the two sets of three \hat{b}_{r_M} portfolios within the extreme $\hat{b}_{\Delta VAR}$ groups. To ensure that the average returns on these portfolio strategies are not influenced by various market-microstructure issues plaguing the smallest stocks, we exclude the five percent of stocks with the lowest *ME* from each cross-section and lag the estimated risk loadings by a month in our sorts.

In the empirical analysis, we consider two main subsamples: early (1931:3-1963:3) and modern (1963:4-2011:4) due to the findings in Campbell and Vuolteenaho (2004) of dramatic differences in the risks of these portfolios between the early and modern period. The first subsample is shorter than that in Campbell and Vuolteenaho (2004) as we require each of the 25 portfolios to have at least one stock as of the time of formation in June.

Finally, we generate a parsimonious cross-section of option, bond, and equity returns for the 1986:1-2011:4 time period based on the findings in Fama and French (1993) and Coval and Shumway (2001). In particular, we use the S&P 100 index straddle returns studied by Coval and Shumway.¹² We also include proxies for the two components of the risky bond factor of Fama and French (1993) which we measure using the return on the Barclays Capital High Yield Bond Index (*HYRET*) and the return on Barclays Capital Investment Grade Bond Index (*IGRET*). When pricing the straddle and risky bond return series, we include the returns on the

¹²Specifically, the series we study includes only those straddle positions where the difference between the options' strike price and the underlying price is between 0 and 5. We thank Josh Coval and Tyler Shumway for providing their updated data series to us.

market (*RMRF*), size (*SMB*), and value (*HML*) equity factors of Fama and French (1993) as they argue these factors do a good job describing the cross-section of average equity returns.

BETA MEASUREMENT

We now examine the validity of an unconditional version of the first-order condition in equation (2.18). We modify equation (2.18) in three ways. First, we use simple expected returns on the left-hand side to make our results easier to compare with previous empirical studies. Second, we condition down equation (2.18) to avoid having to estimate all required conditional moments. Finally, we cosmetically multiply and divide all three covariances by the sample variance of the unexpected log real return on the market portfolio. By doing so, we can express our pricing equation in terms of betas, facilitating comparison to previous research. These modifications result in the following asset-pricing equation

$$E[R_i - R_f] = \gamma \sigma_M^2 \beta_{i,CFM} + \sigma_M^2 \beta_{i,DRM} - \frac{1}{2} \omega \sigma_M^2 \beta_{i,VM}, \quad (2.26)$$

where

$$\begin{aligned} \beta_{i,CFM} &\equiv \frac{Cov(r_{i,t}, N_{CF,t})}{Var(r_{M,t} - E_{t-1}r_{M,t})}, \\ \beta_{i,DRM} &\equiv \frac{Cov(r_{i,t}, -N_{DR,t})}{Var(r_{M,t} - E_{t-1}r_{M,t})}, \\ \text{and } \beta_{i,VM} &\equiv \frac{Cov(r_{i,t}, N_{V,t})}{Var(r_{M,t} - E_{t-1}r_{M,t})}. \end{aligned}$$

We price the average excess returns on our test assets using the unconditional first-order condition in equation (2.26) and the quadratic relationship between the parameters ω and γ given by (2.24). As a first step, we estimate cash-flow, discount-rate, and variance betas using the fitted values of the market's cash flow, discount-rate, and variance news estimated in the previous section. Specifically, we estimate simple WLS regressions of each portfolio's log returns on each news term, weighting each time- $t + 1$ observation pair by the weights used to estimate the VAR in Table 2.1 Panel B. We then scale the regression loadings by the ratio of the sample variance of the news term in question to the sample variance of the unexpected log real return on the market portfolio to generate estimates for our three-beta model.

Characteristic-sorted test assets

Table 2.7 shows the estimated betas for the 25 size- and book-to-market portfolios over the 1931-1963 period. The portfolios are organized in a square matrix with growth stocks at the left, value stocks at the right, small stocks at the top, and large stocks at the bottom. At the right edge of the matrix we report the differences between the extreme growth and extreme value portfolios in each size group; along the bottom of the matrix we report the differences between the extreme small and extreme large portfolios in each BE/ME category. The top matrix displays post-formation cash-flow betas, the middle matrix displays post-formation discount-rate betas, while the bottom matrix displays post-formation variance betas. In square brackets after each beta estimate we report a standard error, calculated conditional on the realizations of the news series from the aggregate VAR model.

In the pre-1963 sample period, value stocks have both higher cash-flow and higher

Table 2.7: Cash-flow, Discount-rate, and Variance Betas in the Early Sample

Panel A: 25 ME- and BE/ME-sorted portfolios

$\hat{\beta}_{CF}$	Growth	2	3	4	Value	Diff
Small	0.44 [0.13]	0.41 [0.11]	0.39 [0.10]	0.41 [0.10]	0.43 [0.10]	-0.01 [0.06]
2	0.29 [0.07]	0.33 [0.09]	0.33 [0.08]	0.36 [0.08]	0.41 [0.10]	0.12 [0.04]
3	0.29 [0.08]	0.27 [0.08]	0.32 [0.09]	0.32 [0.08]	0.44 [0.12]	0.15 [0.05]
4	0.25 [0.07]	0.26 [0.07]	0.30 [0.08]	0.33 [0.08]	0.43 [0.11]	0.18 [0.05]
Large	0.22 [0.07]	0.22 [0.07]	0.25 [0.08]	0.32 [0.10]	0.38 [0.29]	0.16 [0.04]
Diff	-0.22 [0.07]	-0.19 [0.05]	-0.14 [0.04]	-0.09 [0.03]	-0.05 [0.03]	
$\hat{\beta}_{DR}$	Growth	2	3	4	Value	Diff
Small	1.13 [0.15]	1.12 [0.16]	1.09 [0.17]	1.07 [0.17]	1.05 [0.16]	-0.07 [0.07]
2	0.85 [0.11]	0.94 [0.14]	0.91 [0.14]	0.92 [0.16]	1.06 [0.13]	0.21 [0.08]
3	0.86 [0.13]	0.77 [0.09]	0.88 [0.11]	0.86 [0.12]	1.07 [0.15]	0.21 [0.09]
4	0.66 [0.07]	0.75 [0.10]	0.78 [0.09]	0.85 [0.14]	1.11 [0.16]	0.44 [0.13]
Large	0.67 [0.08]	0.63 [0.08]	0.69 [0.11]	0.89 [0.15]	0.89 [0.12]	0.22 [0.13]
Diff	-0.45 [0.14]	-0.49 [0.11]	-0.41 [0.16]	-0.18 [0.13]	-0.17 [0.08]	
$\hat{\beta}_V$	Growth	2	3	4	Value	Diff
Small	-0.32 [0.20]	-0.36 [0.17]	-0.38 [0.18]	-0.37 [0.16]	-0.40 [0.17]	-0.07 [0.09]
2	-0.18 [0.11]	-0.20 [0.14]	-0.25 [0.13]	-0.26 [0.14]	-0.38 [0.18]	-0.20 [0.08]
3	-0.20 [0.13]	-0.13 [0.10]	-0.21 [0.12]	-0.23 [0.13]	-0.39 [0.18]	-0.20 [0.08]
4	-0.04 [0.09]	-0.11 [0.10]	-0.16 [0.12]	-0.24 [0.15]	-0.39 [0.19]	-0.35 [0.13]
Large	-0.05 [0.09]	-0.07 [0.10]	-0.22 [0.16]	-0.31 [0.18]	-0.30 [0.14]	-0.25 [0.11]
Diff	0.27 [0.15]	0.29 [0.10]	0.17 [0.10]	0.05 [0.07]	0.10 [0.07]	

discount-rate betas than growth stocks. An equal-weighted average of the extreme value stocks across size quintiles has a cash-flow beta 0.12 higher than an equal-weighted average of the extreme growth stocks. The difference in estimated discount-rate betas, 0.20, is in the same direction. Similar to value stocks, small stocks have higher cash-flow betas and discount-rate betas than large stocks in this sample (by 0.14 and 0.34, respectively, for an equal-weighted average of the smallest stocks across value quintiles relative to an equal-weighted average of the largest stocks). These differences are extremely similar to those in Campbell and Vuolteenaho (2004), despite the exclusion of the 1929-1931 subperiod, the replacement of the excess log market return with the log real return, and the use of a richer, heteroskedastic VAR.

The new finding in Table 2.7 Panel A is that value stocks and small stocks are also riskier in terms of volatility betas. An equal-weighted average of the extreme value stocks across size quintiles has a volatility beta 0.21 lower than an equal-weighted average of the extreme growth stocks. Similarly, an equal-weighted average of the smallest stocks across value quintiles has a volatility beta that is 0.18 lower than an equal-weighted average of the largest stocks. In summary, value and small stocks were unambiguously riskier than growth and large stocks over the 1931-1963 period.

Table 2.8 reports the corresponding estimates for the post-1963 period. As documented in this subsample by Campbell and Vuolteenaho (2004), value stocks still have slightly higher cash-flow betas than growth stocks, but much lower discount-rate betas. Our new finding here is that value stocks continue to have much lower volatility betas, and the spread in volatility betas is even greater than in the early

Table 2.8: Cash-flow, Discount-rate, and Variance Betas in the Modern Sample

Panel A: 25 ME- and BE/ME-sorted portfolios

$\hat{\beta}_{CF}$	Growth		2		3		4		Value		Diff
Small	0.23	[0.06]	0.24	[0.05]	0.24	[0.04]	0.23	[0.04]	0.26	[0.05]	0.03 [0.03]
2	0.22	[0.05]	0.22	[0.04]	0.24	[0.04]	0.24	[0.04]	0.26	[0.05]	0.04 [0.03]
3	0.20	[0.05]	0.22	[0.04]	0.22	[0.04]	0.23	[0.04]	0.24	[0.04]	0.05 [0.03]
4	0.19	[0.04]	0.21	[0.04]	0.22	[0.04]	0.22	[0.04]	0.24	[0.04]	0.05 [0.03]
Large	0.13	[0.03]	0.17	[0.03]	0.16	[0.03]	0.17	[0.03]	0.19	[0.04]	0.05 [0.03]
Diff	-0.10	[0.04]	-0.07	[0.03]	-0.08	[0.02]	-0.06	[0.02]	-0.07	[0.03]	
$\hat{\beta}_{DR}$	Growth		2		3		4		Value		Diff
Small	1.31	[0.10]	1.06	[0.08]	0.89	[0.07]	0.83	[0.07]	0.87	[0.09]	-0.44 [0.08]
2	1.21	[0.09]	0.97	[0.07]	0.85	[0.06]	0.76	[0.06]	0.80	[0.08]	-0.42 [0.08]
3	1.14	[0.07]	0.89	[0.05]	0.77	[0.06]	0.72	[0.06]	0.72	[0.07]	-0.42 [0.08]
4	1.03	[0.06]	0.85	[0.05]	0.74	[0.06]	0.72	[0.06]	0.75	[0.07]	-0.28 [0.08]
Large	0.84	[0.05]	0.71	[0.04]	0.60	[0.05]	0.59	[0.06]	0.64	[0.06]	-0.20 [0.06]
Diff	-0.46	[0.10]	-0.35	[0.08]	-0.29	[0.06]	-0.24	[0.07]	-0.23	[0.08]	
$\hat{\beta}_V$	Growth		2		3		4		Value		Diff
Small	0.73	[0.29]	0.47	[0.24]	0.34	[0.22]	0.29	[0.20]	0.13	[0.28]	-0.59 [0.11]
2	0.77	[0.27]	0.48	[0.24]	0.32	[0.21]	0.25	[0.22]	0.18	[0.24]	-0.59 [0.10]
3	0.74	[0.25]	0.43	[0.22]	0.32	[0.20]	0.18	[0.22]	0.23	[0.17]	-0.51 [0.12]
4	0.69	[0.23]	0.42	[0.21]	0.24	[0.23]	0.22	[0.24]	0.17	[0.24]	-0.53 [0.10]
Large	0.53	[0.20]	0.41	[0.15]	0.23	[0.17]	0.16	[0.22]	0.17	[0.19]	-0.37 [0.08]
Diff	-0.19	[0.14]	-0.06	[0.12]	-0.11	[0.08]	-0.13	[0.08]	0.03	[0.13]	

period. The volatility beta for the equal-weighted average of the extreme value stocks across size quintiles is 0.52 lower than the volatility beta of an equal-weighted average of the extreme growth stocks, a difference that is more than 42% higher than the corresponding difference in the early period.

One interesting aspect of these findings is the fact that the average β_V of the 25 size- and book-to-market portfolios changes sign from the early to the modern subperiod. Over the 1931-1963 period, the average β_V is -0.25 while over the 1964-2011 period this average becomes 0.36. Of course, given the strong positive link between PE and volatility news documented in the lower right panel of Table 2.5, one should not be surprised that the market's β_V can be positive. Moreover, given the change in sign over time in PE 's correlation with some of the key state variables driving $EVAR$ documented in the Online Appendix, one should not be surprised that β_V changes sign as well.

These results imply that in the post-1963 period where the CAPM has difficulty explaining the low returns on growth stocks relative to value stocks, growth stocks are relative hedges for two key aspects of the investment opportunity set. Consistent with Campbell and Vuolteenaho (2004), growth stocks hedge news about future real stock returns. The novel finding of this paper is that growth stocks also hedge news about the variance of the market return.

Risk-sorted test assets

Table 2.9 shows the estimated betas for the six risk-sorted portfolios over the 1931-1963 period. The portfolios are organized in a rectangular matrix with low CAPM beta stocks at the left, high CAPM beta stocks at the right, low volatility beta

Table 2.9: Betas for Six Risk-Sorted Portfolios in the Early Sample

$\hat{\beta}_{CF}$	Lo \hat{b}_{r_M}		2		Hi \hat{b}_{r_M}		Diff	
Lo \hat{b}_{VAR}	0.22	[0.07]	0.33	[0.09]	0.43	[0.11]	0.21	[0.05]
Hi \hat{b}_{VAR}	0.18	[0.06]	0.26	[0.08]	0.36	[0.10]	0.17	[0.05]
Diff	-0.04	[0.02]	-0.07	[0.03]	-0.08	[0.02]		

$\hat{\beta}_{DR}$	Lo \hat{b}_{r_M}		2		Hi \hat{b}_{r_M}		Diff	
Lo \hat{b}_{VAR}	0.61	[0.07]	0.87	[0.11]	1.09	[0.14]	0.48	[0.09]
Hi \hat{b}_{VAR}	0.55	[0.06]	0.76	[0.09]	0.95	[0.11]	0.40	[0.07]
Diff	-0.07	[0.04]	-0.12	[0.06]	-0.14	[0.05]		

$\hat{\beta}_V$	Lo \hat{b}_{r_M}		2		Hi \hat{b}_{r_M}		Diff	
Lo \hat{b}_{VAR}	-0.08	[0.09]	-0.20	[0.13]	-0.30	[0.16]	-0.22	[0.09]
Hi \hat{b}_{VAR}	-0.06	[0.09]	-0.09	[0.10]	-0.17	[0.13]	-0.11	[0.07]
Diff	0.02	[0.05]	0.11	[0.07]	0.13	[0.06]		

stocks at the top, and high volatility beta stocks at the bottom. At the right edge of the matrix we report the differences between the high CAPM beta and the low CAPM beta portfolios in each volatility beta group; along the bottom of the matrix we report the differences between the high volatility beta and the low volatility beta portfolios in each CAPM beta category. As in Panel A, the top matrix displays post-formation cash-flow betas, the middle matrix displays post-formation discount-rate betas, while the bottom matrix displays post-formation volatility betas.

In the pre-1963 sample period, high CAPM beta stocks have both higher cash-flow and higher discount-rate betas than low CAPM beta stocks. An equal-weighted average of the high CAPM beta stocks across the two volatility beta categories has a cash-flow beta 0.19 higher than an equal-weighted average of the low CAPM beta stocks. The difference in estimated discount-rate betas is 0.44 and in the same

direction. Similar to high CAPM beta stocks, low volatility beta stocks have higher cash-flow betas and discount-rate betas than high volatility beta stocks in this subsample (by 0.06 and 0.11, respectively, for an equal-weighted average of the low volatility beta stocks across the three CAPM beta categories relative to a corresponding equal-weighted average of the high volatility beta stocks).

High CAPM beta stocks and low volatility beta stocks are also riskier in terms of volatility betas. An equal-weighted average of the high CAPM beta stocks across volatility beta categories has a post-formation volatility beta 0.16 lower than an equal-weighted average of the low CAPM beta stocks. Similarly, an equal-weighted average of the low volatility beta stocks across CAPM beta categories has a post-formation volatility beta that is 0.09 lower than an equal-weighted average of the high volatility beta stocks. In summary, high CAPM beta and low volatility beta stocks were unambiguously riskier than low CAPM beta and high volatility beta stocks over the 1931-1963 period.

Table 2.10 shows the estimated betas for the six risk-sorted portfolios over the post-1963 period. In the modern period, high CAPM beta stocks again have higher cash-flow and higher discount-rate betas than low CAPM beta stocks. An equal-weighted average of the high CAPM beta stocks across the two volatility beta categories has a cash-flow beta 0.08 higher than an equal-weighted average of the low CAPM beta stocks. The difference in estimated discount-rate betas is 0.55 and in the same direction. However, high CAPM beta stocks are no longer riskier in terms of volatility betas. Now, an equal-weighted average of the high CAPM beta stocks across the two volatility beta categories has a post-formation variance beta 0.28 higher than

Table 2.10: Betas for Six Risk-Sorted Portfolios in the Modern Sample

\widehat{b}_{VAR} Lo	0.16	[0.03]	0.17	[0.03]	0.25	[0.05]	0.08	[0.04]
\widehat{b}_{VAR} Hi	0.15	[0.03]	0.17	[0.04]	0.23	[0.05]	0.08	[0.04]
Diff	-0.01	[0.02]	0.00	[0.02]	-0.01	[0.02]		

$\widehat{\beta}_{DR}$	Lo \widehat{b}_{r_M}		2	Hi \widehat{b}_{r_M}		Diff	
\widehat{b}_{VAR} Lo	0.55	[0.05]	0.71	[0.05]	1.11	[0.09]	0.56 [0.08]
\widehat{b}_{VAR} Hi	0.73	[0.06]	0.95	[0.06]	1.27	[0.09]	0.54 [0.11]
Diff	0.18	[0.07]	0.24	[0.07]	0.16	[0.06]	

$\widehat{\beta}_V$	Lo \widehat{b}_{r_M}		2	Hi \widehat{b}_{r_M}		Diff	
\widehat{b}_{VAR} Lo	0.22	[0.19]	0.31	[0.22]	0.50	[0.29]	0.27 [0.13]
\widehat{b}_{VAR} Hi	0.44	[0.16]	0.64	[0.18]	0.72	[0.27]	0.28 [0.15]
Diff	0.21	[0.07]	0.33	[0.09]	0.22	[0.06]	

a corresponding equal-weighted average of the low CAPM beta stocks. Since, in the three-beta model, covariation with aggregate volatility has a negative premium, the three-beta model can potentially explain why stocks with high past CAPM betas have offered relatively little extra return, at least in the modern period.

In the post-1963 period, sorts on volatility beta continue to generate economically and statistically significant spread in post-formation volatility beta. An equal-weighted average of low volatility beta stocks across the three CAPM beta categories has a post-formation volatility beta that is 0.26 lower than the post-formation volatility beta of a corresponding equal-weighted average of high volatility beta stocks. Sorts on volatility beta also generate spread in discount-rate beta, but essentially no spread in cash-flow betas in the post-1963 period.

Non-equity test assets

Finally, Table 2.11 reports the three ICAPM betas of the S&P 100 index straddle position analyzed in Coval and Shumway (2001) along with the corresponding ICAPM betas of the three equity factors and the default bond factor of Fama and French (1993) over the period 1986:1 - 2011:4. Consistent with the nature of a straddle bet, we find that the straddle has a very large volatility beta of 1.51 along with a large negative discount-rate beta of -1.71 and a large (relatively speaking) negative cash-flow beta of -0.39. As one would expect, the betas of the Fama-French equity factors are consistent with the findings for the size- and book-to-market-sorted portfolios in Table 2.8 Panel B. Finally, the riskier component of Fama and French's (1993) risky bond factor, *HYRET*, has a cash-flow beta of 0.06, a discount-rate beta of 0.26, and a volatility beta of -0.20. These betas are economically and statistically significant from those of the safer component, *IGRET*. The difference in volatility beta between *HYRET* and *IGRET* is consistent with the fact that risky corporate debt is short the option to default.

BETA PRICING

We next turn to pricing the cross-section with these three ICAPM betas. We evaluate the performance of five asset-pricing models: 1) the traditional CAPM that restricts cash-flow and discount-rate betas to have the same price of risk and sets the price of variance risk equal to zero; 2) the two-beta intertemporal asset pricing model of Campbell and Vuolteenaho (2004) that restricts the price of discount-rate risk to equal the variance of the market return, 3) our three-beta intertemporal asset pricing model that restricts the price of discount-rate risk to equal the variance of the market

Table 2.11: Betas for Option, equity, and bond portfolios

Panel C: Option, equity, and bond portfolios						
	<i>STRADDLE</i>	<i>RMRF</i>	<i>SMB</i>	<i>HML</i>	<i>HYRET</i>	<i>IGRET</i>
$\hat{\beta}_{CF}$	-0.39 [0.28]	0.18 [0.05]	0.04 [0.02]	0.05 [0.03]	0.06 [0.02]	0.00 [0.01]
$\hat{\beta}_{DR}$	-1.71 [0.46]	0.81 [0.06]	0.19 [0.05]	-0.26 [0.09]	0.26 [0.07]	0.03 [0.03]
$\hat{\beta}_V$	1.51 [0.86]	-0.02 [0.29]	-0.01 [0.07]	-0.47 [0.11]	-0.20 [0.21]	0.05 [0.03]

return and constrains the price of cash-flow and variance risk to be related by equation (2.24), with $\rho = 0.95$ per year; 4) a partially-constrained three-beta model that restricts the price of discount-rate risk to equal the variance of the market return but freely estimates the other two risk prices (effectively decoupling γ and ω), and 5) an unrestricted three-beta model that allows free risk prices for cash-flow, discount-rate, and volatility betas. Each model is estimated in two different forms: one with a restricted zero-beta rate equal to the Treasury-bill rate, and one with an unrestricted zero-beta rate following Black (1972).

Characteristic-sorted test assets

Table 2.12 reports results for the early sample period 1931-1963, using 25 size- and book-to-market-sorted portfolios as test assets. The table has ten columns, two specifications for each of our five asset pricing models. The first 16 rows of Table 2.12 are divided into four sets of four rows. The first set of four rows corresponds to the zero-beta rate (in excess of the Treasury-bill rate), the second set to the premium on cash-flow beta, the third set to the premium on discount-rate beta, and the fourth set to the premium on volatility beta. Within each set, the first row reports the point estimate in fractions per quarter, and the second row annualizes this estimate, multiplying by 400 to aid in interpretation. These parameters are estimated from a cross-sectional regression

$$\bar{R}_i^e = g_0 + g_1 \hat{\beta}_{i,CF_M} + g_2 \hat{\beta}_{i,DR_M} + g_3 \hat{\beta}_{i,V_M} + e_i, \quad (2.27)$$

where a bar denotes time-series mean and $\bar{R}_i^e \equiv \bar{R}_i - \bar{R}_{rf}$ denotes the sample average simple excess return on asset i . The third and fourth rows present two alternative

standard errors of the monthly estimate, described below.

Below the premia estimates, we report the R^2 statistic for a cross-sectional regression of average returns on our test assets onto the fitted values from the model. We also report a composite pricing error, computed as a quadratic form of the pricing errors. The weighting matrix in the quadratic form is a diagonal matrix with the inverse of the sample test asset return volatilities on the main diagonal.

Standard errors are produced with a bootstrap from 10,000 simulated realizations. Our bootstrap experiment samples test-asset returns and first-stage VAR errors, and uses the first-stage and second-stage WLS VAR estimates in Table 2.1 to generate the state-variable data.¹³ We partition the VAR errors and test-asset returns into two groups, one for 1931 to 1963 and another for 1963 to 2011, which enables us to use the same simulated realizations in subperiod analyses. The first set of standard errors (labeled A) conditions on estimated news terms and generates betas and return premia separately for each simulated realization, while the second set (labeled B) also estimates the first-stage and second-stage VAR and the news terms separately for each simulated realization. Standard errors B thus incorporate the considerable additional sampling uncertainty due to the fact that the news terms as well as betas are generated regressors.

Two alternative 5-percent critical values for the composite pricing error are produced with a bootstrap method similar to the one we have described above, except that the test-asset returns are adjusted to be consistent with the pricing model before the random samples are generated. Critical values A condition on estimated

¹³When simulating the bootstrap, we drop realizations which would result in negative RVAR and redraw.

news terms, while critical values B take account of the fact that news terms must be estimated.

Finally, Table 2.12 reports the implied risk-aversion coefficient, γ , which can be recovered as g_1/g_2 , as well as the sensitivity of news about risk to news about market variance, ω , which can be recovered as $-2 * g_3/g_2$. The three-beta ICAPM estimates are constrained so that both γ and the implied ω are strictly positive.

Table 2.12 shows that in the 1931-1963 period, the restricted three-beta model explains the cross-section of stock returns reasonably well. The cross-sectional R^2 statistics are almost 56% for both forms of this model. Both the Sharpe-Lintner and Black versions of the CAPM do a slightly poorer job describing the cross-section (both R^2 statistics are roughly 52%). The two-beta ICAPM of Campbell and Vuolteenaho (2004) performs slightly better than the CAPM and slightly worse than the volatility ICAPM. None of the theoretically-motivated models considered are rejected by the data based on the composite pricing test. Consistent with the claim that the three-beta model does a good job describing the cross-section, Table 2.12 shows that the constrained and the unrestricted factor model barely improve pricing relative to the three-beta ICAPM.

Figure 2.6 provides a visual summary of these results. The figure plots the predicted average excess return on the horizontal axis and the actual sample average excess return on the vertical axis. In summary, we find that the three-beta ICAPM improves pricing relative to both the Sharpe-Lintner and Black versions of the CAPM.

Table 2.12: Asset Pricing Tests for the Early Sample

Parameter	CAPM	2-beta ICAPM	3-beta ICAPM	Constrained	Unrestricted
\hat{R}_{zb} less R_f (\hat{g}_0)	0	-0.002	0	0.002	0
% per annum	0%	-0.90%	0%	0.82%	0%
Std. err.	0	(0.016)	0	(0.015)	0
$\hat{\beta}_{CF}$ premium (\hat{g}_1)	0.038	0.040	0.096	0.082	0.075
% per annum	15.11%	15.82%	38.33%	32.91%	29.88%
Std. err.	(0.015)	(0.024)	(0.145)	(0.091)	(0.129)
$\hat{\beta}_{DR}$ premium (\hat{g}_2)	0.038	0.040	0.016	0.016	0.018
% per annum	15.11%	15.82%	6.40%	6.40%	7.28%
Std. err.	(0.015)	(0.024)	(0.004)	(0.004)	(0.074)
$\hat{\beta}_V$ premium (\hat{g}_3)			-0.012	-0.019	-0.020
% per annum			-4.73%	-7.67%	-7.99%
Std. err.			(0.063)	(0.166)	(0.193)
\hat{R}^2	53.01%	53.12%	54.72%	55.75%	55.75%
Pricing error	0.024	0.023	0.022	0.021	0.020
5% critic. val.	(0.064)	(0.031)	(0.106)	(0.053)	(0.049)
Implied $\hat{\gamma}$	N/A	N/A	5.99	5.14	N/A
Implied $\hat{\omega}$	N/A	N/A	N/A	1.28	N/A

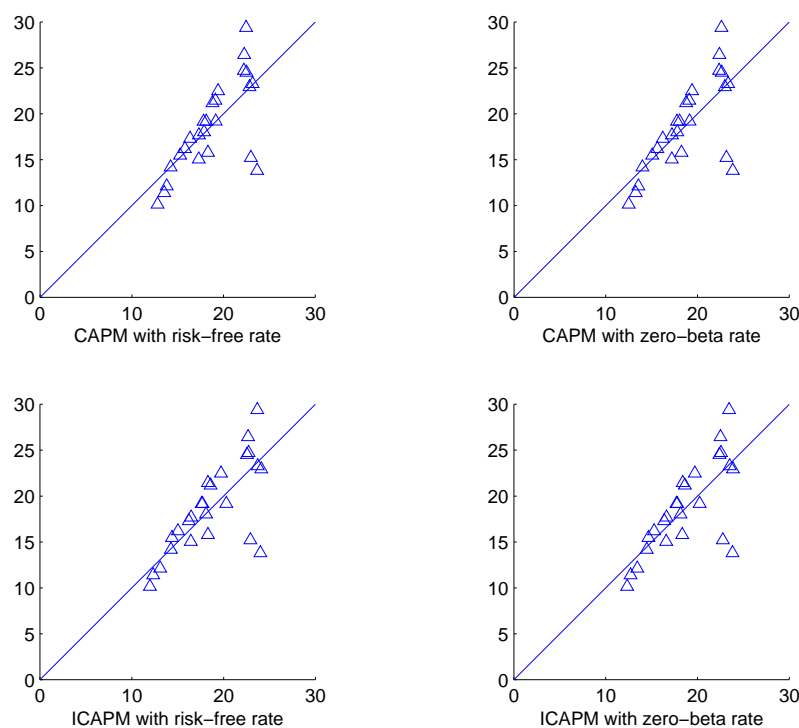


Figure 2.6: Pricing 25 Size and Value Portfolios, Early Period

The four diagrams correspond to (clockwise from the top left) the CAPM with a constrained zero-beta rate, the CAPM with an unconstrained zero-beta rate, the three-factor ICAPM with a free zero-beta rate, and the three-factor ICAPM with the zero-beta rate constrained to the risk-free rate. The horizontal axes correspond to the predicted average excess returns and the vertical axes to the sample average realized excess returns for the 25 ME- and BE/ME-sorted portfolios. The predicted values are from regressions presented in Table 2.12 for the sample period 1931:3-1963:2.

This success is due in part to the inclusion of volatility betas in the specification. For the Black version of the three-beta ICAPM, the spread in volatility betas across the 25 size- and book-to-market-sorted portfolios generates an annualized spread in average returns of 1.46% compared to a comparable spread of 7.41% and 3.18% for cash-flow and discount-rate betas. Variation in volatility betas accounts for 2% of the variation in explained returns compared to 39% and 7% for cash-flow and discount-rate betas respectively. The remaining 52% of the explained variation in average returns is due of course to the covariation among the three types of betas.

Results are very different in the 1963-2011 period. Table 2.13 shows that in this period, both versions of the CAPM do a very poor job of explaining cross-sectional variation in average returns on portfolios sorted by size and book-to-market. When the zero-beta rate is left as a free parameter, the cross-sectional regression picks a negative premium for the CAPM beta and implies an R^2 of roughly 5%. When the zero-beta rate is constrained to the risk-free rate, the CAPM R^2 falls to roughly -37%. Both versions of the static CAPM are easily rejected at the five-percent level by both sets of critical values.

In the modern period, the unconstrained zero-beta rate version of the two-beta Campbell and Vuolteenaho (2004) model does a better job describing the cross-section of average returns than the CAPM. However, the implied coefficient of risk aversion, 20.70, is arguably extreme.

The three-beta model with the restricted zero-beta rate also does a poor job explaining cross-sectional variation in average returns across our test assets. However, if we continue to restrict the risk price for discount-rate and variance news but allow

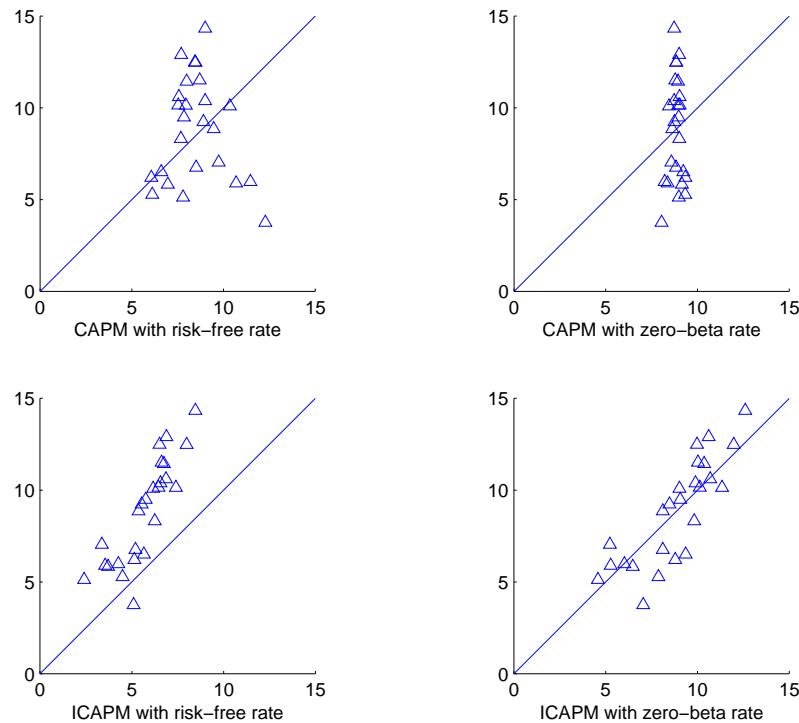


Figure 2.7: Pricing 25 Size and Value Portfolios, Modern Period

The four diagrams correspond to (clockwise from the top left) the CAPM with a constrained zero-beta rate, the CAPM with an unconstrained zero-beta rate, the three-factor ICAPM with a free zero-beta rate, and the three-factor ICAPM with the zero-beta rate constrained to the risk-free rate. The horizontal axes correspond to the predicted average excess returns and the vertical axes to the sample average realized excess returns for the 25 ME- and BE/ME-sorted portfolios. The predicted values are from regressions presented in Table 2.13 for the sample period 1963:3-2011:4.

an unrestricted zero-beta rate, the explained variation increases to roughly 69%, three-quarters larger than the R^2 of the corresponding two-beta ICAPM. The estimated risk price for cash-flow beta is an economically reasonable 30 percent per year with an implied coefficient of relative risk aversion of 9.63. Both versions of our intertemporal CAPM with stochastic volatility are not rejected at the 5-percent level by either set of critical values.

Figure 2.7 provides a visual summary of these results. For the Black version of the

three-beta ICAPM, spread in volatility betas across the 25 size- and book-to-market-sorted portfolios generates an annualized spread in average returns of 6.52% compared to a comparable spread of 3.90% and 2.24% for cash-flow and discount-rate betas. Variation in volatility betas accounts for 92% of the variation in explained returns compared to 20% for cash-flow betas as well as 7% for discount-rate betas. Covariation among the three types of betas is responsible for the remaining -19% of explained variation in average returns.

The relatively poor performance of the risk-free rate version of the three-beta ICAPM is due to the derived link between γ and ω . To show this, Figure 2.8 provides two contour plots (one each for the risk-free and zero-beta rate versions of the model in the top and bottom panels of the figure respectively) of the R^2 resulting from combinations of (γ, ω) ranging from (0,0) to (40,16). On the same figure we also plot the relation between γ and ω derived in equation (2.24). The top panel of Figure 2.8 shows that even with the intercept restricted to zero, R^2 's are as high as 70% for some combinations of (γ, ω) . Unfortunately, as the plot shows, these combinations do not coincide with the curve implied by equation (2.24). Once the zero-beta rate is unconstrained, the contours for R^2 's greater than 60% cover a much larger area of the plot and coincide nicely with the ICAPM relation of equation (2.24).

Consistent with the contour plots of Figure 2.8, the pricing results in Table 2.13 based on the partially-constrained factor model further confirms that the link between γ and ω is responsible for the poor fit of the restricted zero-beta rate version of the three-beta ICAPM in the modern period. When removing the constraint linking γ and ω but leaving the constraint on the discount-rate beta premium in place,

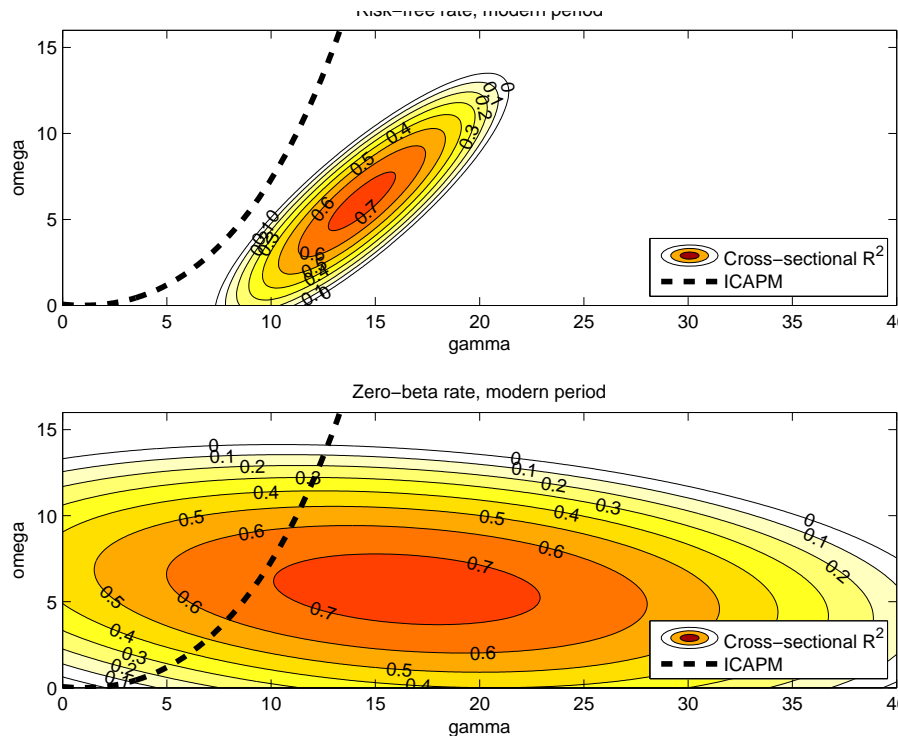


Figure 2.8: Contour Plots Showing Goodness-of-Fit

The two contour plots show how the R^2 of the cross-sectional regression explaining the average returns on the 25 size- and book-to-market portfolios varies for different values of γ and ω for the risk-free rate (top panel) and zero-beta rate (bottom panel) three-beta ICAPM model estimated in Table 2.13 for the sample period 1963:3-2011:4. The two plots also indicate the approximate ICAPM relation between γ and ω described in equation (2.24).

the R^2 increases from -57% to 74%. Nevertheless, the risk prices for γ and ω remain economically large and of the right sign.

Risk-sorted test assets

We confirm that the success of the three-beta ICAPM is robust by expanding the set of test portfolios beyond the 25 size- and book-to-market-sorted portfolios. First, we show that our three-beta model not only describes the cross-section of characteristics-sorted portfolios but also can explain the average returns on risk-sorted portfolios. We examine risk-sorted portfolios as Daniel and Titman (1997, 2012) and Lewellen, Nagel, and Shanken (2010) argue that asset-pricing tests using only portfolios sorted by characteristics known to be related to average returns, such as size and value, can be misleading due to the low-dimensional factor structure of the 25 size and book-to-market-sorted portfolios.

Table 2.14 prices the six risk-sorted portfolios described in Table 2.7 Panel B in conjunction with six of the 25 size- and book-to-market-sorted portfolios of Table 2.7 Panel A (the low, medium, and high BE/ME portfolios within the small and large ME quintiles). We continue to find that the three-beta ICAPM improves pricing relative to both the Sharpe-Lintner and Black versions of the CAPM. Moreover, the relatively high R^2 (57%) is not disproportionately due to characteristics-sorted portfolios as the R^2 for the risk-sorted subset (69%) is not only comparable to but also larger than the R^2 for the characteristics-sorted subset (51%). Figure 2.9 shows this success graphically.

Table 2.15 prices the cross-section of characteristics- and risk-sorted portfolios in

Table 2.14: Early Sample Asset Pricing Tests with Risk-sorted Portfolios

Parameter	CAPM	2-beta ICAPM	3-beta ICAPM	Constrained	Unrestricted
\hat{R}_{zb} less R_f (\hat{g}_0)	0	0.002	0	0.003	0
% per annum	0%	0.73%	0%	1.20%	0%
Std. err.	0	(0.015)	0	(0.013)	0
$\hat{\beta}_{CF}$ premium (\hat{g}_1)	0.035	0.034	0.085	0.072	0.035
% per annum	14.05%	13.47%	34.05%	28.82%	-30.06%
Std. err.	(0.015)	(0.023)	(0.141)	(0.089)	(0.179)
$\hat{\beta}_{DR}$ premium (\hat{g}_2)	0.035	0.034	0.016	0.016	0.010
% per annum	14.05%	13.47%	6.40%	6.40%	24.18%
Std. err.	(0.015)	(0.023)	(0.004)	(0.004)	(0.124)
$\hat{\beta}_V$ premium (\hat{g}_3)			-0.009	-0.007	-0.100
% per annum			-3.69%	-2.86%	-25.65%
Std. err.			(0.046)	(0.037)	(0.258)
\hat{R}^2	54.08%	54.17%	54.58%	56.82%	68.95%
characteristics	46.98%	46.77%	48.95%	51.15%	68.43%
risk-sorted	73.65%	73.14%	65.68%	68.78%	61.61%
Pricing error	0.012	0.012	0.012	0.011	0.012
5% critic. val.	(0.042)	(0.018)	(0.060)	(0.031)	(0.023)
Implied $\hat{\gamma}$	N/A	N/A	5.32	4.50	N/A
Implied $\hat{\omega}$	N/A	N/A	N/A	0.89	N/A

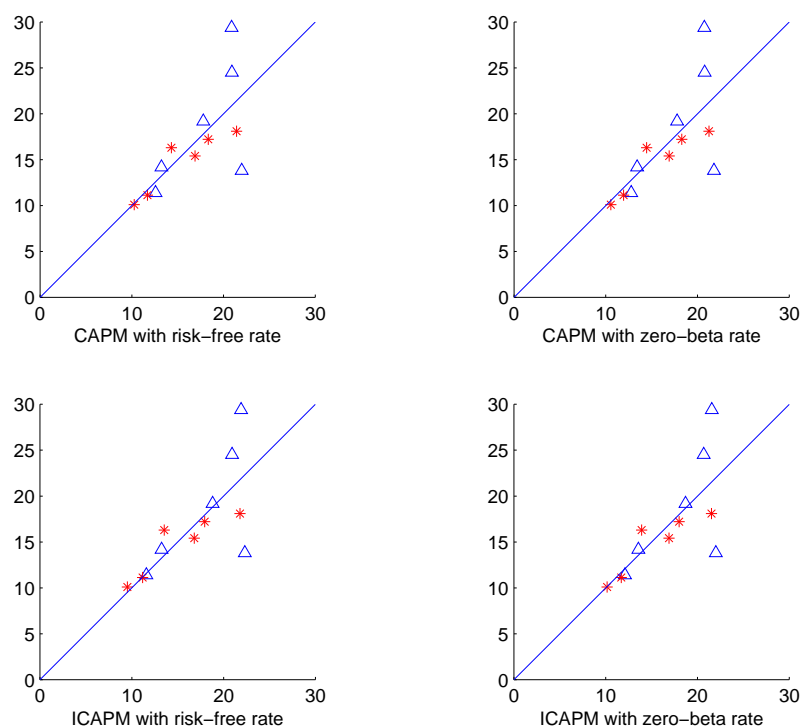


Figure 2.9: Pricing Risk Sorted Portfolios, Early Period

The four diagrams correspond to (clockwise from the top left) the CAPM with a constrained zero-beta rate, the CAPM with an unconstrained zero-beta rate, the three-factor ICAPM with a free zero-beta rate, and the three-factor ICAPM with the zero-beta rate constrained to the risk-free rate. The horizontal axes correspond to the predicted average excess returns and the vertical axes to the sample average realized excess returns for the 25 ME- and BE/ME-sorted portfolios. The predicted values are from regressions presented in Table 2.13 for the sample period 1963:3-2011:4.

the modern period. We find that the zero-beta rate three-beta ICAPM is not rejected by the data while both versions of the CAPM are rejected. Again, the relatively high R^2 for the zero-beta rate version of the volatility ICAPM (76%) is not disproportionately due to characteristics-sorted portfolios as the R^2 for the risk-sorted subset (81%) is not only comparable to but also larger than the R^2 for the characteristics-sorted subset (77%). Figure 2.10 provides a graphically summary of these results.

Non-equity test assets

We also show that our three-beta model can help explain average returns on non-equity portfolios designed to be highly correlated with aggregate volatility risk, namely the S&P 100 index straddles of Coval and Shumway (2001). We first calculate the expected return on straddle portfolio based on the estimates of the zero-beta rate volatility ICAPM in Table 2.13. The contributions to expected quarterly return from the straddle's cash-flow, discount-rate, and volatility betas are -2.92%, -1.33%, and -3.87% respectively. As the average quarterly realized return on the straddle is -21.66%, an equity-based estimate of the three-beta model explains roughly 38% of the realized straddle premium.

Table ?? shows that our intertemporal CAPM with stochastic volatility is not rejected at the 5-percent level when we price the joint cross-section of equity, bond, and straddle returns. The implied risk aversion coefficient (roughly 15 for both the risk-free and zero-beta rate implementations of the model) is high but not unreasonable. In sharp contrast, the CAPM is strongly rejected. Though the two-beta

Table 2.15: Modern Sample Asset Pricing Tests with Risk-sorted Portfolios

Parameter	CAPM	2-beta ICAPM	3-beta ICAPM	Constrained	Unrestricted
\hat{R}_{zb} less R_f (\hat{g}_0)	0	0.017	0	0.008	0
% per annum	0%	6.60%	0%	2.98%	0%
Std. err.	0	(0.009)	0	(0.010)	0
$\hat{\beta}_{CF}$ premium (\hat{g}_1)	0.016	0.001	0.057	0.078	0.136
% per annum	6.26%	0.44%	22.63%	31.28%	54.41%
Std. err.	(0.009)	(0.014)	(0.086)	(0.102)	(0.153)
$\hat{\beta}_{DR}$ premium (\hat{g}_2)	0.016	0.001	0.008	0.008	-0.001
% per annum	6.26%	0.44%	3.11%	3.11%	-0.58%
Std. err.	(0.009)	(0.014)	(0.002)	(0.002)	(0.063)
$\hat{\beta}_V$ premium (\hat{g}_3)			-0.015	-0.027	-0.019
% per annum			-6.15%	-10.77%	-7.61%
Std. err.			(0.025)	(0.105)	(0.123)
\hat{R}^2	-19.85%	8.49%	16.45%	75.92%	74.97%
characteristics	-4.98%	15.02%	28.30%	76.66%	77.21%
risk-sorted	-48.82%	10.69%	-11.29%	81.42%	81.42%
Pricing error	0.050	0.041	0.032	0.016	0.016
5% critic. val.	(0.035)	(0.022)	(0.049)	(0.102)	(0.040)
Implied $\hat{\gamma}$	N/A	N/A	7.28	N/A	N/A
Implied $\hat{\omega}$	N/A	N/A	3.96	N/A	N/A

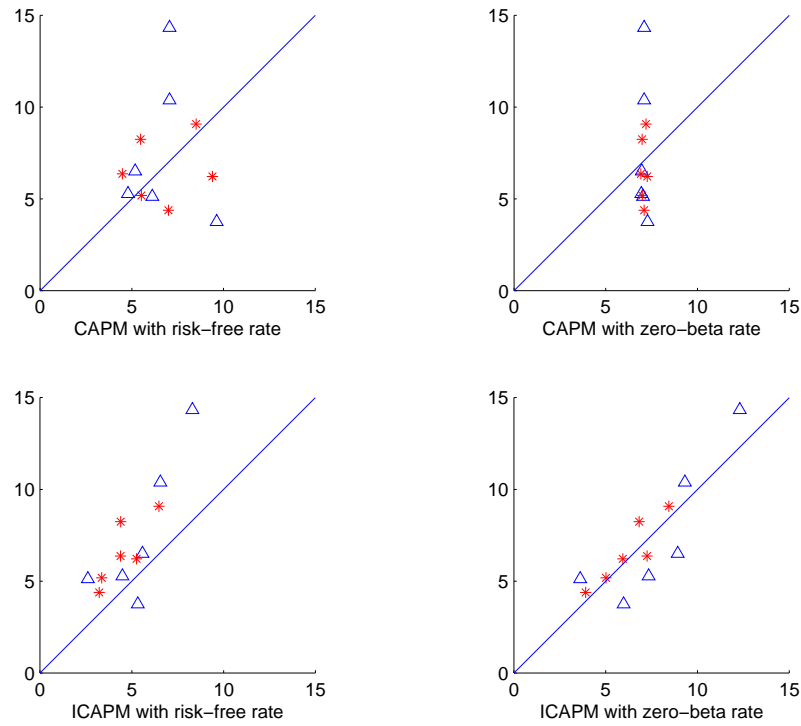


Figure 2.10: Pricing Risk Sorted Portfolios, Modern Period

The four diagrams correspond to (clockwise from the top left) the CAPM with a constrained zero-beta rate, the CAPM with an unconstrained zero-beta rate, the three-factor ICAPM with a free zero-beta rate, and the three-factor ICAPM with the zero-beta rate constrained to the risk-free rate. The horizontal axes correspond to the predicted average excess returns and the vertical axes to the sample average realized excess returns for six ME- and BE/ME-sorted portfolios (denoted by triangles) and six risk-sorted portfolios (denoted by asterisks). The predicted values are from regressions presented in Table 2.14 for the sample period 1931:3-1963:2.

ICAPM is not rejected, the required risk aversion is too extreme (over 53 for both versions of the model) to be realistic.

Summary of US financial history

Figure 2.11 (third panel) plots the time-series of the smoothed combined shock $\gamma N_{CF} - N_{DR} - \frac{1}{2}\omega N_V$ based on the estimate of the zero-beta model for the modern period (Table 2.13). The correlation of this shock with the associated N_{CF} is 0.90. Similarly, the correlation of this shock with the associated N_{DR} is 0.26. Finally, the correlation of this shock with the associated N_V is -0.76. Figure 2.11 also plots the corresponding smoothed shock series for the CAPM ($N_{CF} - N_{DR}$) and for the two-beta ICAPM ($\gamma N_{CF} - N_{DR}$). The two-beta model shifts the history of good and bad times relative to the CAPM, as emphasized by Campbell, Giglio, and Polk (2012). The model with stochastic volatility further accentuates that periods with high market volatility, such as the 1930s and the late 2000s, are particularly hard times for long-term investors.

2.4 THE CONTRIBUTION OF STOCHASTIC VOLATILITY

We extend the approximate closed-form intertemporal capital asset pricing model of Campbell (1993) to allow for stochastic volatility. Our model recognizes that an investor's investment opportunities may deteriorate either because expected stock returns decline or because the volatility of stock returns increases. A conservative long-term investor will wish to hedge against both types of changes in investment opportunities; thus, a stock's risk is determined not only by its beta with unexpected

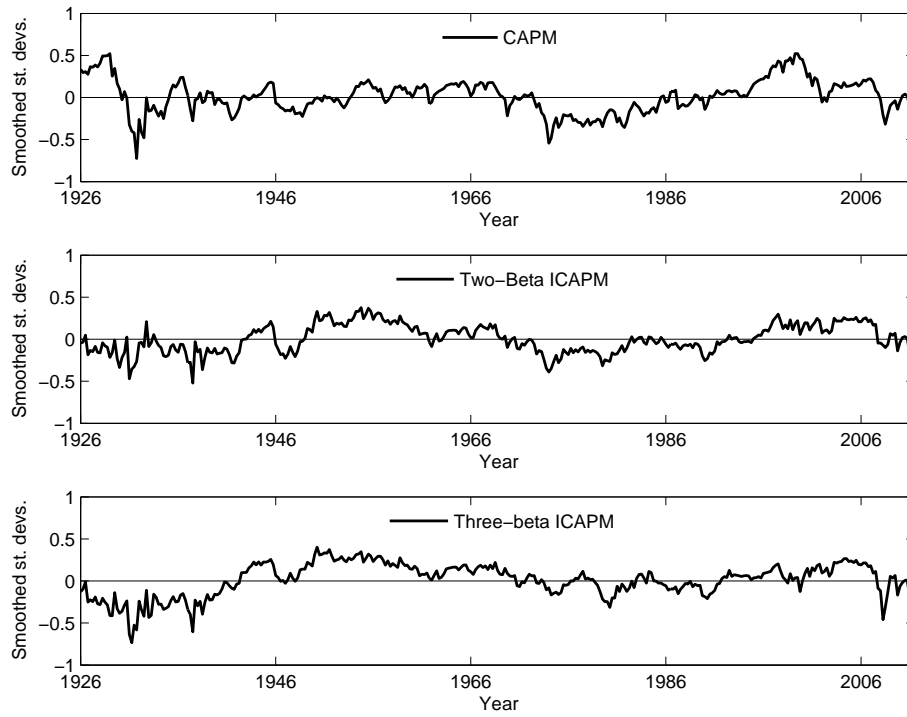


Figure 2.11: Pricing Risk Sorted Portfolios, Modern Period

This figure plots the time-series of the smoothed combined shock for the CAPM ($N_{CF} - N_{DR}$), the two-beta ICAPM ($\gamma N_{CF} - N_{DR}$), and the three-beta ICAPM that includes stochastic volatility ($\gamma N_{CF} - N_{DR} - \frac{1}{2} \omega N_V$) for the unconstrained zero-beta rate specifications estimated in Table 8 for the modern subperiod. The shock is smoothed with a trailing exponentially-weighted moving average. The decay parameter is set to 0.08 per quarter, and the smoothed news series is generated as $MA_t(SDF) = 0.08SDF_t + (1 - 0.08)MA_{t-1}(N)$. This decay parameter implies a half-life of six years. The sample period is 1926:2-2011:4.

market returns and news about future returns (or equivalently, news about market cash flows and discount rates), but also by its beta with news about future market volatility. Although our model has three dimensions of risk, the prices of all these risks are determined by a single free parameter, the coefficient of relative risk aversion.

Our implementation models the return on the aggregate stock market as one element of a vector autoregressive (VAR) system; the volatility of all shocks to the VAR is another element of the system. The empirical implementation of our VAR reveals new low-frequency movements in market volatility tied to the default spread. We show that the negative post-1963 CAPM alphas of growth stocks are justified because these stocks hedge long-term investors against both declining expected stock returns, and increasing volatility. The addition of volatility risk to the model helps it to deliver a moderate, economically reasonable value of risk aversion.

Our empirical work is limited in one important respect. We test only the unconditional implications of the model and do not evaluate its conditional implications. A full conditional test is likely to be a challenging hurdle for the model. To see why, recall that we assume a rational long-term investor always holds 100% of his or her assets in equities. However, time-variation in real stock returns generally gives the long-term investor an incentive to shift the relative weights on cash and equity, unless real interest rates and market volatility move in exactly the right way to make the equity premium proportional to market volatility. Although we do not explicitly test whether this is the case, previous work by Campbell (1987) and Harvey (1989, 1991) rejects this proportionality restriction.

One way to support the assumption of constant 100% equity investment is to

invoke binding leverage constraints. Indeed, in the modern sample, the Black (1972) version of our three-beta model is consistent with this interpretation as the estimated difference between the zero-beta and risk-free rates is positive, statistically significant, and economically large. However, the risk aversion coefficient we estimate may be too large to explain why leverage constraints should bind.

Nevertheless, our model does directly answer the interesting microeconomic question: Are there reasonable preference parameters that would make a long-term investor, constrained to invest 100% in equity, content to hold the market rather than tilting towards value stocks or other high-return stock portfolios? Our answer is clearly yes.

The high brokerage charges and the heavy transfer tax...sufficiently diminish the liquidity of the market. But a little consideration of this expedient brings us up against a dilemma, and shows how the liquidity of investment markets often facilitates, though it sometimes impedes, the course of new investment.

John Maynard Keynes

3

Informative Prices and the Cost of Capital Markets

INVESTORS SPEND A GREAT DEAL of time and money speculating on financial valuations or hiring others to trade on their behalf. While criticizing speculation is always fashionable, the scale of the recent increase in resources spent on capital markets has many people concerned that we are wasting talent and resources. There seems to be little consensus among financial economists regarding the value of this speculative activity; however, it is easy to observe the increase in quantity. Historically, the share of national income spent on financial market activity remained

relatively stable until the mid-1970s, when the financial sector began to grow much more rapidly than the aggregate US economy. Before rushing to judge whether we now spend too much, or too little, on active investing, we need theory and evidence that promise to explain the root cause of this growth and the resulting effect on asset prices.

In this paper, I document how the sharp decline in the cost of financial transactions facilitated the modern increase in financial activity. To clarify the forces at work, I present a stylized model of an economy with a financial sector that allows investors to trade ownership claims on a risky investment. The supply of investment responds to asset prices, and investor demand drives costly financial activity. Investors decide how much of their resources to employ researching the future prospects of the uncertain outcome, and market transaction costs affect the quantity and time horizon of informed speculation. We see the surprising result that the financial sector consumes more resources through spending on active investing as it operates more efficiently. As dynamic trading strategies become feasible, the model suggests that the information content of asset prices increases, especially over short-horizons.

Historical data on US market activity and asset prices confirm these predictions. The most significant decrease in transaction costs occurred in 1975, when on May Day the SEC demanded that stock exchanges end the practice of forcing a fixed commission schedule on all equity transactions. In response to broker competition, the average cost of institutional trading plummeted to about half of previous levels.¹ This event is significant not only in the historical time series, but it also provides a

¹US Securities and Exchange Commission, Directorate of Economic and Policy Research. *Staff Report on the Securities Industry in 1978* (July 26, 1979)

natural setting for identifying the causal mechanism. This regulatory change leads to a surge in capital market spending, trading, and compensation, with an impact that predictably varies across investment characteristics and time horizons.

The efficiency of modern financial markets enables dynamic trading strategies and encourages investors to spend more resources on research and trading, but increased efficiency does not necessarily align the incentives of private speculators toward activities with the greatest social benefit. Returning again to the stylized model shows that increases in the efficiency of financial market operations may lead to less efficient economic outcomes.

SPENDING ON CAPITAL MARKET ACTIVITY

Consider how much the United States spends on capital market activities each year as a share of total national production. Figure 3.1 shows the cost of capital markets as a percentage of the GDP of the US private sector, where capital market spending consists of the profits and employee compensation tabulated using the gross value added measures reported by United States Bureau of Economic Analysis (BEA).² The cost of capital markets is remarkably stable for approximately half a century. Beginning with a cost of 0.27% of GDP in 1920 to a cost of 0.35% in 1970, spending stays fairly close to its average of 0.32% with the exception of a moderate dip around World War II. Then, a little before 1980, we notice a dramatic surge in the cost of capital markets to the point where capital markets now consume two percent of annual spending.

²A complete description of the underlying data will be available in an online appendix.

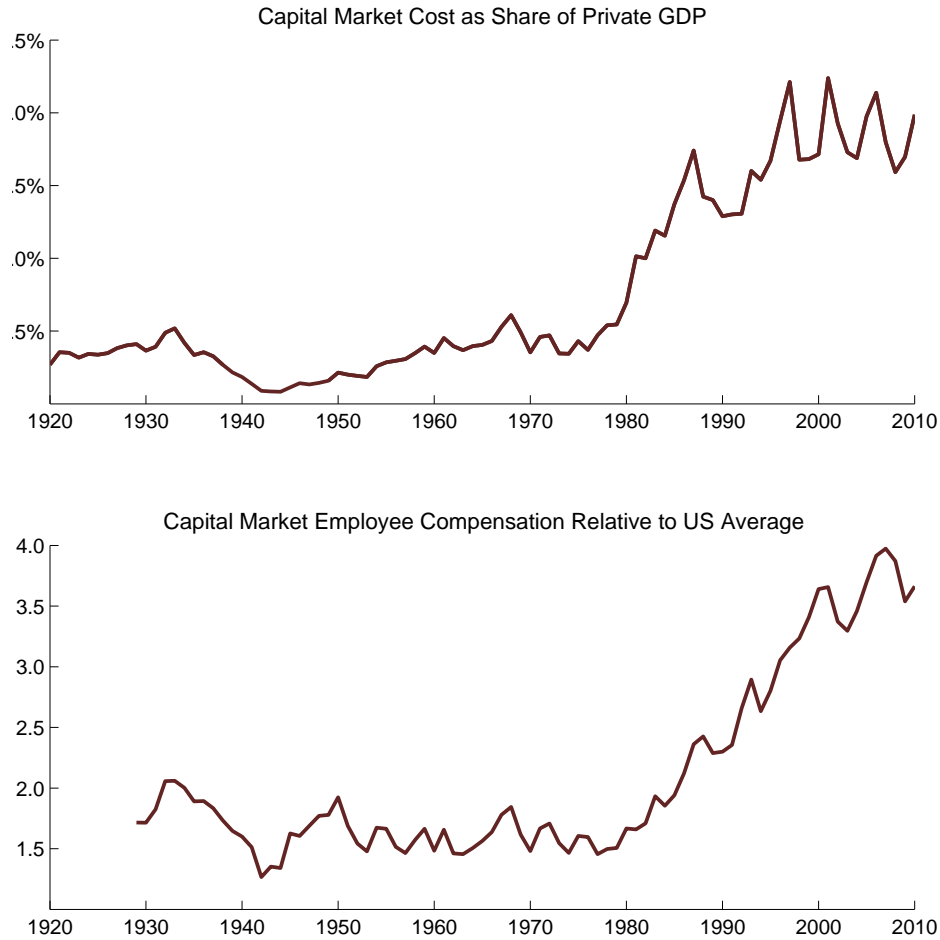


Figure 3.1: Capital Market Spending and Compensation

The upper plot shows the share of GDP attributed to the capital markets sector using the gross value added measure, and the lower plot shows the ratio of average employee compensation in the capital markets sector relative to the US private industry average. The primary source for these calculations is the industry accounts data published by the US Bureau of Economic Analysis as of March 2011. Capital markets-related industries are described in Table 3.1. Data prior to 1947 comes from Philippon (2012).

Philippon (2012) lays out the scope of the historical challenge as he tabulates the costs and quantities of various financial activities over the past 130 years in the United States. In his analysis, it appears that the unit cost of financial intermediation has remained relatively stable over time despite advancements in technology. He notes a puzzling increase in the cost of financial activity over the past 30 years that he cannot explain with a corresponding increase in the quantity or quality of financial services.

With a particular focus on this modern period, Greenwood and Scharfstein (2012) attribute the modern growth of the financial sector as a whole to two specific components: an increase in active investing and an expansion in credit markets. To contrast these two culprits, I allocate the corresponding financial activities from the national industry accounts data, as shown in Table 3.1. The resources consumed in credit and banking activities grew significantly over the past century but follow a distinct pattern from the resources spent investing in financial markets. The upper plot in Figure 3.2 shows both activities consumed a growing fraction of GDP, but the cost of banking and credit expanded at steady consistent pace since World War II while the surge in trading and investing seems to be a more recent phenomenon. Unlike the capital markets sector, the lower plot of Figure 3.2 shows the historical compensation of employees in the banking and credit sector differs only slightly from the private sector average and increases only moderately in recent decades.

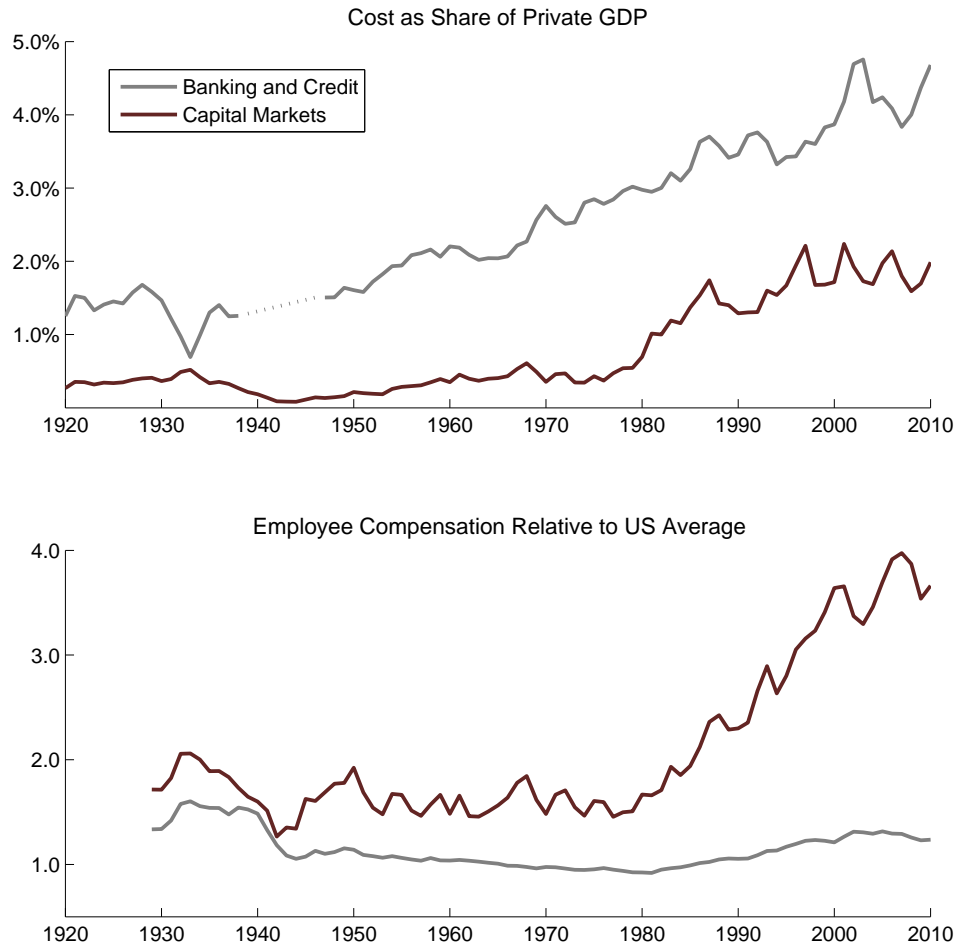


Figure 3.2: Contrasting Banking and Credit vs. Capital Market Activities

The upper plot contrasts the cost of banking and credit activity with the cost of capital markets using gross value added, and the lower plot shows the respective employee compensation ratios relative to the US private industry average. The primary source for these calculations is the industry accounts data published by the US Bureau of Economic Analysis as of March 2011. The classification to industry groups is shown in Table 3.1. Data prior to 1947 comes from Philippon (2012).

Table 3.1: Financial sector components in national income accounts

This table shows the components of the financial sector and the associated NAICS codes as used by the US Bureau of Economic Analysis in their national income accounts. The grouping of the components has not always been historically consistent. The highlighted industries are those which will be termed the capital markets sector and are the primary focus of this paper.

Finance, Insurance, and Real Estate
Banking and Credit (521 & 522)
Banking
Credit agencies other than banks
Capital Markets (523 & 525)
Security and commodity brokers
Funds, trusts, and other financial vehicles
Holding and other investment offices
Insurance (524)
Insurance carriers
Insurance agents, brokers, and service
Real Estate and Leasing (531, 532, 533)
Real Estate
Rental and leasing services and lessors of intangible assets

THEORIES OF FINANCIAL INVESTMENT DISTORTIONS

Dissatisfaction with the quantity of talent and resources consumed by financial markets seems to peak during economic downturns. Amidst the Great Depression, Keynes criticized American financial markets, arguing, “when the capital development of a country becomes the by-product of the activities of a casino, the job is likely to be ill-done.”³ On the other hand, the broad impact of financial crises could also suggest we need a large and highly compensated financial sector to replace animal spirits with dispassionate analysts.

Certainly, there is a need to understand the circumstances and incentives that pull resources toward financial markets. What gives rise to a distorted financial sector?

³Keynes, John Maynard, *The General Theory of Employment, Interest and Money* (London: Macmillan, 1936), page 159.

Economic research offers three explanations for outsized financial activity: irrational investors do not know they trade too much, rational investors cannot help trading too much, or perhaps the industry is rife with rent-seeking.

Financial markets seem to be amazingly adroit at exploiting irrational beliefs and behaviors. Fanciful trading or the decision to pay exorbitant fees to popular investment managers may funnel unnecessary fees into finance and have other negative consequences (De Long, Shleifer, Summers and Waldmann, 1989).

In a model where market participants are assumed to be rational, they may still spend too much on active investment because inference is difficult (Pástor and Stambaugh, 2010) or out of a desire to avoid being the greater fool when negotiating transactions. Glode, Green and Lowery (2012) present this situation as an arms race externality for financial expertise. The model presented by Bolton, Santos and Scheinkman (2011) has a similar mechanism; opaque markets attract talent and more informed valuations lure the best investments away from public exchanges.

These explanations capture important aspects of financial markets, but neither seems uniquely modern. If traders are foolish now, they were foolish before. Shrewd traders will always prefer to be better informed than their counterparty. We are forced to ask: what changed?

Philippon and Reshef (2013) point toward the rent-seeking channel, and propose the growth in compensation is a result of deregulation. The active government oversight intended to curb the worst excesses in the financial markets of the 1920s was gradually relaxed 50 years later, and Philippon and Reshef propose rents lured talent from more productive endeavors (Murphy, Shleifer and Vishny, 1991).

Supporting this view, Bai, Philippon and Savov (2012) suggest modern asset prices show no increase in their information content over the past 50 years. They suggest the increase in financial spending may result from rent extraction, suggesting the growth in active investment has had little effect on asset prices.

UNDERSTANDING THE CAUSES AND CONSEQUENCES OF THE COST OF CAPITAL MARKETS

With so much highly compensated talent flowing into investment management, it is hard to believe that asset prices are no more informative in the modern information age than they were in the bygone era when investors in top hats exchanged small pieces of paper. As an alternative explanation for the root cause of the modern growth of capital markets, I propose technological efficiency. The decreasing cost of transacting makes dynamic trading strategies feasible and draws talent and technology toward acquiring faster paced information. Confirming the results of Bai et al. (2012), I find only very weak evidence that modern asset prices capture more long-horizon information; however, I find strong evidence of an increase in active trading and information content at horizons of less than one year.

To help frame the empirical findings, I present a stylized model illustrating the role of trading horizons in costly capital markets. The key comparative static will measure the effect of increases in trading efficiency. The model predicts that as the cost of financial activity decreases, total spending in the financial sector actually increases, especially for short-horizon speculation.

This explanation has a large degree of empirical success in explaining aggregate

spending on capital markets over time, particularly in regard to aggregate spending on active investing (French, 2008). More efficient transaction costs lead to higher quantities of informed trading, providing an underlying explanation for Greenwood and Scharfstein's observation that the observed growth of modern finance coincides with a growth in actively investing. The events of May 1975 highlight the significance of this mechanism, as the SEC instituted rule 19-b and replaced the high trading commissions enforced by stock exchange members with competitive transaction rates. Using this event and information from historical fee schedules, we observe how the operational efficiency of capital markets affects the financial industry and market prices.

This paper provides new evidence on the changes that caused and accompanied the modern growth in the cost of capital markets. Linking these findings to economic theory clarifies the underlying incentives and opens the door to the broader question of whether the returns to finance are worth the cost.

3.1 A STYLIZED MODEL OF CAPITAL MARKETS

In this section, I present a stylized model of capital markets where the supply of the risky investment responds to asset prices and where the financial market is costly to operate. I will show how changes in the cost of transacting affect the quantity of resources spent on finance and affect the characteristics of asset prices.

To better understand the role financial markets play, consider an illustrative, general equilibrium framework where investors spend resources in acquiring information and engaging in costly transactions. In the spirit of the Q-theory

(Brainard and Tobin, 1977), the supply of investment will respond to the market price, so the information in asset prices plays a key role in capital allocation.

Ultimately, we want to observe how changes in the cost of transacting affects the resources spent in capital markets. Additionally, the model will distinguish between short-run and long-run behavior, generating novel predictions relating the growth in capital market spending to asset prices which will be confirmed in the data.

Unlike the opaque bilateral setting of Glode et al. (2012), all market prices in the model will be publicly observed, which has historically been true for equity markets and is becoming increasingly common across asset classes. The setup more closely resembles the endogenous information setting of Grossman and Stiglitz (1980), adding the salient features necessary to model a costly financial market and multiple time horizons.

The key comparative statics will be the impact of an exogenous change of transaction costs on total capital market spending and the information content of asset prices, noting the differential impact by trading horizon. I briefly mention the welfare implications in section 3.4.

THE SETTING

THE SUPPLY OF RISKY INVESTMENT

Consider a risky investment traded publicly over a T periods ($t \in [1, T]$) prior to yielding an uncertain payout X consumer in period $T + 1$, where the uncertain

component of X is

$$X - E[X] = \sum_{t=1}^T \theta_t + \varepsilon. \quad (3.1)$$

Each of the component random variables are independent, mean-zero, and normally distributed with variances σ_θ^2 and σ_ε^2 . The full, random component $\sum \theta_t$ becomes public knowledge in period $T + 1$. However, market participants can spend resources to discover the information in period t , and they will be termed long-horizon investors. Alternately, short-horizon investors may spend a smaller amount of resources to discover each piece of short horizon information (θ_t) in period t . The random component ε cannot be observed prior to period $T + 1$.

The quantity of the risky investment is responsive to investment demand, allowing the quantity of shares in one period, Q_t , to increase or decrease with the market price, P_t . For simplicity, we'll model this as a linear supply curve, with slope parameter $b > 0$. The change in investment supply will be

$$Q_{t+1} - Q_t = b (P_t - P_{t-1}). \quad (3.2)$$

where the initial price is assumed to be the unconditional expectation, $P_0 = E_0[P_1]$. By construction, the supply of investment is fixed in the short-run (contemporaneous with the trading period) and responds to financial market prices over longer horizons (the next period).

INVESTORS AND FINANCIAL MARKETS

The agents will be modeled by a continuum of identical investors. Wealth can be transferred across periods at an interest rate of zero and is consumed in the final period. Each investor is endowed with w_o units of wealth (measured in units of final consumption) and a share, q_o , of the risky investment. By construction, the total initial quantity of investment is $Q_o = \int_{i \in [0,1]} q_{o,i} di$.

Individuals can choose whether they want to acquire information and actively speculate based on the difference between their valuation and the observed market price. To learn the full value of $\sum \theta_t$ during the first trading period requires paying k_L , whereas short-horizon traders who only learn each component θ_t at time t pay $k_S \leq k_L$. Alternately, investors may choose to infer their valuations from the public market price. Since their valuations will not differ from the market price, they will not actively trade and I'll refer to these traders as passive, though they might make trades driven by changes in their uncertainty.

Each individual seeks to maximize expected CARA utility of final consumption. For convenience, we'll denote the consumption of investor i as their final wealth, w_i , with associated expected utility $E[-\exp\{-aw_i\}]$ for absolute risk aversion parameter a .

Investors must commit whether to spend resources on information in period $t = 0$ before any trading happens. In subsequent periods prior to the final outcome, investors may choose to trade their holdings of the risky asset at the prevailing market price. The transaction costs associated with capital markets are passed directly through to investors. For analytical convenience, we'll assume they take a quadratic

form so that the trading from a prior holding of $q_{i,t-1}$ shares in period $t - 1$ to $q_{i,t}$ during the trading in period t will result in a transaction cost of $\frac{c}{2} (q_t - q_{t-1})^2$.

We can describe the evolution of investor wealth as

$$w_{i,t+1} = w_{i,t} + q_{i,t} (P_{t+1} - P_t) - \frac{c}{2} (q_{i,t+1} - q_{i,t})^2 \quad (3.3)$$

where agents are identically endowed with w_0 consumption and q_0 shares of the risky investment. In the final period, the price of the risky investment will simply be the outcome, i.e. $P_{T+1} = X$.

PORTFOLIO CHOICE

The linear-CARA-normal framework allows the expected utility from the perspective of investor i in trading period t to be calculated as

$$E_{i,t} [-\exp \{-aw_i\}] = -\exp \left\{ -aE_{i,t} [w_i] + \frac{a^2}{2} \text{Var}_{i,t} [w_i] \right\}. \quad (3.4)$$

Through monotonic transformations, the investor can maximize the certainty-equivalent, which takes the mean-variance form, $E_{i,t} [w_i] - \frac{a}{2} \text{Var}_{i,t} [w_i]$. The concavity of the problem suggests we can find the optimal portfolio in each period, $q_{i,t}^*$, at the point where the first order condition holds,

$$\frac{\partial}{\partial q_{i,t}} E_{i,t} [w_{i,3}] = \frac{a}{2} \frac{\partial}{\partial q_{i,t}} \text{Var}_{i,t} [w_{i,3}].$$

To motivate the optimal portfolio rules, we can work backwards from the final trading period. The optimal portfolio $q_{i,T}^*$ in last trading period that maximizes the

utility of consumption in the subsequent period will have the associated first order condition

$$q_{i,T}^* = \frac{E_{i,T}[X - P_T] + cq_{i,T-1}}{a\text{Var}_{i,T}[X] + c}. \quad (3.5)$$

This is the classic myopic portfolio rule with a transaction cost adjustment. In the numerator, we see the optimal portfolio increases linearly with the expected return, $E_{i,T}[X - P_T]$. The second term in the numerator shows how much transaction costs discourage trading by anchoring the portfolio at the initial position, $q_{i,T-1}$. The magnitude of the transaction costs, c , determines the extent to which this affects the optimal portfolio.

In solving the model, I will show how the anchoring feature of transaction costs results in optimal portfolio rules that are a weighted average of their myopic, one-period expected return and the returns offered in future periods.

EQUILIBRIUM

In this setting, investors can be grouped into three types based on their information sets. The mass of agents of type j are those who pay k_j for their investment information will be measured as the quantity $\lambda_j \in [0, 1]$.

Definition In a rational expectations equilibrium,

- (a) markets will clear
- (b) investors will choose to spend resources on information to maximize ex ante utility, leading to an allocation $\{\lambda_L, \lambda_S\}$ and where $\lambda_N = 1 - \lambda_L - \lambda_S$ is the fraction of individuals who will only infer information from market prices

- (c) investors of each type have an optimal demand function $q_{i,t}(P_t)$ for the risky asset conditional on the market price, which will be constructed from their rational beliefs about random variables $(\theta_t$ and v_t) conditional on the observed price.

MARKET CLEARING

It will be useful to explicitly define market clearing. Noisy supply shocks will add uncertainty so that the market price does not perfectly reveal all information. Specifically, the total quantity of investment supply will equal investment demand,

$$Q_t = \sum_i \lambda_i q_{i,t} + \frac{v_t}{a\sigma_\varepsilon^2 + c}, \quad (3.6)$$

comprising the sum of the individual demands $(q_{i,t})$ times the mass of the investor type (λ_i) plus the scaled demand shock $v_t \sim N(\sigma_v^2)$. The values in the denominator scale the shock by variance and transaction costs. In this sense, the noise can be interpreted in the same way as the demands of an informed investor, as can be seen from demand function (3.5), but obviously the shock is unrelated to the actual final payout of the investment.

INTUITION

To build the intuition behind this model and its equilibrium, consider Figure 3.3. For this particular illustration, this will assume just one trading period ($T = 1$) and there is no distinction between long-horizon and short-horizon informed investors, though

the paper will generally consider $T > 1$ in order to highlight the importance of time horizon. The left panel plots the fraction of informed speculators along the horizontal axis, ranging from 0 to 1. The vertical axis measures expected utility for both the informed speculators and the expected utility for the uninformed, passive investors. When there are no informed speculators, the information advantage is obvious as the expected utility for informed active investors is significantly higher than that of the passive investors who observe only the market price. As the fraction of the informed investors increases, the difference between the two expected utilities decreases. This is the general case, and the intuition extends to the multiple period setting; as the market price becomes more informative the relative advantage of paying for the information decreases. With these parameters, the equilibrium point of indifference between acquiring the costly information occurs at the point where approximately $\frac{1}{4}$ of the investors acquire the costly information. To the right of the equilibrium point, the trading profits resulting from learning more about the risky outcome θ are not worth the resources it could cost (k).

On the right panel, the horizontal axis continues to measure the fraction of informed speculators, and on the vertical axis we see the equilibrium price. In the case of no informed investors, the variation in price is entirely due to the supply shocks v . As the fraction of informed traders increases, we see two effects. The average price increases as investors are willing to commit more capital to investment because there is less uncertainty. Additionally, the variance of the market prices increases. This is because the price now also contains information about the investment prospects. Not surprisingly, the information content of asset prices levels

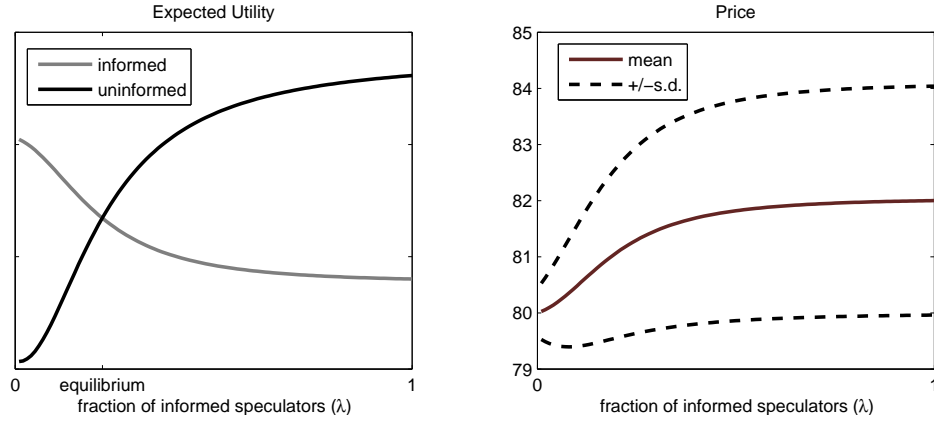


Figure 3.3: Intuition behind model equilibrium

The plots above correspond to the model presented in the paper in the one-period setting, $T = 1$. The model parameters are: $Q_0 = 1$, $\bar{X} = 100$, $\sigma_\varepsilon^2 = \sigma_\theta^2 = 10^2$, $\sigma_v^2 = 2^2$, $a = 0.1$, $c = 10$, and $k = 1$. For illustration, the investment supply is allowed to be elastic in the short-run ($\Delta Q = b(P - E[P])$), with linear supply parameter $b = 0.2$. The left axis plots the expected utility for the informed speculators and the uninformed passive investors. The right axis plots how the distribution of the market price, P , changes with respect to the quantity of informed speculators.

of around the equilibrium point, further evidence that little additional value is gained acquiring information that is already largely in the market price.

Proposition 1 (Equilibria) *There exist rational expectations equilibria under the assumed parameter restrictions ($0 < k_S < k_L$).*

The proof for the one-period case ($T = 1$) should be clear from the discussion above. There will be no long-horizon traders. Since the expected utilities are continuous in $\lambda \in [0, 1]$, we simply need to appeal to the intermediate value theorem for existence. The difference between the expected utility of the informed and uninformed traders will nearly always be monotonically decreasing in λ , which guarantees uniqueness.

The same intermediate value approach guarantees a unique solution in the case of multiple periods ($T > 2$) in the case where one or more type is always inferior and has optimal weight zero. The existence of the multiple horizon solution when there is a positive mass of each of the three types can be motivated by working backwards from the final period. In the final period, informed traders face a situation identical to the one-period model. In prior periods, the relative advantage to the long-horizon information is decreasing in λ_L . The mass of investors in λ_S will be uninformed about the information θ_{t+k} (for $k > 1$), and like the uninformed investors, can infer more information as λ_L increases. As long as there are positive quantities of each investor type, the marginal effect of more traders will follow the same relative rank impact on ex ante utility, guaranteeing a unique solution.

CHARACTERIZING A MULTIPLE HORIZON SOLUTION ($T = 2$)

To characterize the analytical differences between long-horizon and short-horizon speculation, I will more fully characterize the solution for $T = 2$. In this setting, the outcome will be a long-run event in the first period and a short-run event in the second period, which immediately precedes the investment outcome. After this short-horizon trading is complete, investor i will consume

$$w_i = w_o + q_o P_1 + q_{i,1} (P_2 - P_1) + q_{i,2} (X - P_2) - \frac{c}{2} ((q_{i,1} - q_o)^2 + (q_{i,2} - q_{i,1})^2) - k_i. \quad (3.7)$$

ASSUMING LINEARITY AND THE RESULTING EXPECTATIONS

To calculate the investor demand functions, we need to know their expectations, which will be affected by the information they perceive from the market prices they observe. I will assert and then prove that the market prices can be expressed as linear functions of the unknown variables,

$$P_1 = \bar{P}_1 + \beta_1 \theta_1 + \beta_2 \theta_2 + \beta_{v_1} v_1 \quad (3.8)$$

and

$$P_2 = \bar{P}_2 + \beta_p (P_1 - \bar{P}) + \beta_3 \theta_1 + \beta_4 \theta_2 + \beta_{v_2} v_2. \quad (3.9)$$

The unknown coefficients are derived in the appendix, thus confirming the assumed linear functional form.

Additionally, to help with the notation and intuition, we note that the beliefs of uninformed and short-run traders hold about X from observing the market price in period 1 will be affected by the variation in price. We can express these expectations as

$$E_{S,1} [X] = \bar{X} + \rho_{S,1} Y_{S,1} \quad (3.10)$$

where

$$Y_{S,1} = \theta_2 + \frac{\beta_{v_1}}{\beta_2} v_1 \quad (3.11)$$

$$\propto (P_1 - \bar{P}_1 - \beta_1 \theta_1) \quad (3.12)$$

and so that $\rho \in [0, 1]$ is a simple function of the assumed parameters

$$\rho = \frac{\sigma_{\theta}^2}{\sigma_{\theta}^2 + \left(\frac{\beta_{v_{1,1}}}{\beta_{\theta,1}} \right)^2 \sigma_v^2}.$$

The investors who have spent no resources on information simply take valuations from their deviation from the market price

$$(E_{N,1} [X] - \bar{X}) \propto (P_1 - \bar{P}) \quad (3.13)$$

PORTFOLIO OPTIMIZATION IN PERIOD 2

The investors will be categorized by the trading period in which they receive information about θ : in the long-horizon (L), short-horizon (S) and not at all (N).

For each of the three investor types (L , S , and N), we can express their optimal portfolio in terms of their prior position and their current expectations $E_{i,2} [X]$ and $\text{Var}_{i,2} [X]$. The long-run and short-run speculators will both know θ_1 and θ_2 in period 2 so $E_{L,2} [X] = E_{S,2} [X]$. The associated variance will be $\text{Var}_{L,2} [X] = \text{Var}_{S,2} [X] = \sigma_{\varepsilon}^2$. From (3.5) we can conclude that the optimal portfolio for these two types of investors will be

$$q_{L,2}^* = \frac{(\bar{X} + \theta_1 + \theta_2 - P_2) + cq_{L,1}^*}{a\sigma_{\varepsilon}^2 + c} \quad (3.14)$$

and

$$q_{S,2}^* = \frac{(\bar{X} + \theta_1 + \theta_2 - P_2) + cq_{S,1}^*}{a\sigma_{\varepsilon}^2 + c} \quad (3.15)$$

The optimal portfolio for the investors who purchase no information

$$q_{N,2}^* = \frac{E_{N,2} [X - P_2] + cq_{N,1}^*}{a\text{Var}_{N,2} [X] + c} \quad (3.16)$$

depends on the expectations, $E_{n,2} [\theta]$ and $\text{Var}_{n,2} [\theta]$, which will be derived later.

PORTFOLIO OPTIMIZATION IN PERIOD 1

When investing for the long-run (in period 1), investors choose their allocation aware of their optimal short-run portfolio rules in equations (3.14 - 3.16). Those short-run rules show that each portfolio allocation is linearly related to the expected return ($E_i [X - P_2]$) and the prior portfolio allocation ($q_{i,1}$).

The form of the period 1 demand function for long-horizon investors is similar to that of the other two investor types. It is derived by substituting the period 1 demand from equation (3.14) into equation (3.7) and taking the first order conditions to find the optimal portfolio

$$q_{L,1}^* = \frac{(1 - \Gamma) E_{L,1} [P_2 - P_1] + \Gamma E_{L,1} [X - P_1] + cq_o}{\Omega + c \left(1 + \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)} \quad (3.17)$$

where the tilt toward the long-run return is

$$\Gamma = \underbrace{\frac{c}{a\sigma_\varepsilon^2 + c}}_{\text{return next period}} + a \underbrace{\frac{(2a\sigma_\varepsilon^2 + c) \beta_{v_2}^2 \sigma_v^2 a^2 \sigma_\varepsilon^4}{(a\sigma_\varepsilon^2 + c)^4}}_{\text{prefer to avoid adverse } v_2}$$

and the variance

$$\Omega = \underbrace{\left(\frac{c}{a\sigma_\varepsilon^2 + c} \right)^2 \sigma_\varepsilon^2}_{\text{variance of } X} + \underbrace{\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^4 \beta_{v_2,2}^2 \sigma_v^2}_{\text{variance in } P_2}.$$

To develop some intuition for this long-horizon portfolio rule in equation (3.17), consider the three terms in the numerator. As before, there is a weight pulling the optimal portfolio toward the initial position, q_0 as a result of transaction costs. The other two terms are a weighted average of the myopic expected return, $E_{L,1}[P_2 - P_1]$ and the long-run expected return, $E_{L,1}[X - P_1]$, with respective weights $(1 - \Gamma)$ and Γ .

The weight Γ that the investor tilts toward the long-horizon return will always be weakly positive, $\Gamma \in [0, 1)$, and its magnitude will increase with transaction costs. The relationship with transaction costs arises from the investor recognizing positions taken today will persist into the future due to the anchoring effect of transaction costs. Additionally, there is some uncertainty in the price next period, so investors have an incentive to lock in P_1 now rather than pay an uncertain P_2 .

The demand functions for the short-run and uninformed investors take an identical form, with slightly different values for Γ and Ω .

DERIVING INVESTOR DEMAND

This section derives the demand functions for the model with two trading periods ($T = 2$). For each investor, we use their expectations to maximize the utility of final

wealth, as defined in equation (3.7),

$$w_i = w_o - k_i + q_o P_1 + q_{i,1} (P_2 - P_1) + q_{i,2} (X - P_2) - \frac{c}{2} ((q_{i,1} - q_o)^2 + (q_{i,2} - q_{i,1})^2).$$

The first order condition, $\frac{\partial}{\partial q_{i,t}} E_{i,t} [w_i] = \frac{a}{2} \frac{\partial}{\partial q_{i,t}} \text{Var}_{i,t} [w_i]$, can be used to derive the investor demand functions. In period 2, the only source of uncertainty is X and we get

$$q_{i,2}^* = \frac{E_{i,2} [X - P_2] + cq_{i,1}}{a\text{Var}_{i,2} [X] + c},$$

which leads to the optimal demand functions presented for each type of investor, as in (3.5).

Deriving the demand functions for period 1 with multiple horizons requires a fair amount of algebra. Beginning with the expression for expected wealth,

$$\begin{aligned} E_{i,1} [w_i] &= w_o - k_i + q_o P_1 + q_{i,1} E_{i,1} [P_2 - P_1] + E_{i,1} [q_{i,2} (X - P_2)] \\ &\quad - \frac{c}{2} ((q_{i,1} - q_o)^2 + E_{i,1} [(q_{i,2} - q_{i,1})^2]), \end{aligned}$$

we can substitute in period 2's demand function

$$\begin{aligned} E_{i,1} [w_i] &= w_o + q_o P_1 + q_{i,1} E_{i,1} [P_2 - P_1] + E_{i,1} \left[\frac{E_{i,2} [X - P_2] + cq_{i,1}}{a\text{Var}_{i,2} [X] + c} (X - P_2) \right] \\ &\quad - \frac{c}{2} \left((q_{i,1} - q_o)^2 + E_{i,1} \left[\left(\frac{E_{i,2} [X] - P_2 + cq_{i,1}}{a\text{Var}_{i,2} [X] + c} - q_{i,1} \right)^2 \right] \right) \end{aligned}$$

with first derivative

$$\begin{aligned} \frac{\partial}{\partial q_{i,1}} E[w_i] &= E_{i,1}[P_2 - P_1] + \frac{c E_{i,1}[X - P_2]}{a \text{Var}_{i,2}[X] + c} - c(q_{i,1} - q_o) \\ &\quad - c \left(\frac{a \text{Var}_{i,2}[X]}{a \text{Var}_{i,2}[X] + c} \right)^2 q_{i,1} + c \frac{a \text{Var}_{i,2}[X] E_{i,1}[X - P_2]}{(a \text{Var}_{i,2}[X] + c)^2} \end{aligned}$$

so the final expression is

$$\begin{aligned} \frac{\partial}{\partial q_{i,1}} E[w_i] &= E_{i,1}[P_2 - P_1] \\ &\quad + \left(\frac{c}{a \text{Var}_{i,2}[X] + c} + c \frac{a \text{Var}_{i,2}[X]}{(a \text{Var}_{i,2}[X] + c)^2} \right) E_{i,1}[X - P_2] \\ &\quad + c q_o - c \left(1 + \left(\frac{a \text{Var}_{i,2}[X]}{a \text{Var}_{i,2}[X] + c} \right)^2 \right) q_{i,1} \end{aligned}$$

The optimal portfolio in period one will be the one that solves the first order condition,

$$q_{i,1}^* = \frac{E_{i,1}[P_2 - P_1] + \left(\frac{c}{a \text{Var}_{i,2}[X] + c} + c \frac{a \text{Var}_{i,2}[X]}{(a \text{Var}_{i,2}[X] + c)^2} \right) E_{i,1}[X - P_2] + c q_o}{\frac{a}{2q_{i,1}^*} \text{Var}_{i,1}[w_i] + \left(1 + \left(\frac{a \text{Var}_{i,2}[X]}{a \text{Var}_{i,2}[X] + c} \right)^2 \right) c}.$$

The expected values for P_2 and X are apparent from the assumed linearity in (3.8) and (3.9), so the task at hand is to come up with expressions for $\frac{a}{2q_{i,1}^*} \text{Var}_{i,1}[w_3]$, where

the variance term can be expressed as

$$\begin{aligned}
\text{Var}_{i,1} [w_i] &= \text{Var}_{i,1} \left[q_{i,1} P_2 + q_{i,2} (X - P_2) - \frac{c}{2} (q_{i,2} - q_{i,1})^2 \right] \\
&= \text{Var}_{i,1} \left[q_{i,1} P_2 + q_{i,2} (X - P_2) - \frac{c}{2} q_{i,2}^2 + c q_{i,1} q_{i,2} \right] \\
&= \text{Var}_{i,1} \left[q_{i,1} P_2 + q_{i,2} (X - E_{i,2} [X]) + q_{i,2}^2 \left(a \text{Var}_{i,2} [X] + \frac{c}{2} \right) \right]
\end{aligned}$$

and the remaining calculation requires using the expectations of each investor and calculating the sensitivity with respect to the first period allocation.

LONG-HORIZON INVESTORS IN PERIOD 1

For long-horizon investors, the uncertain terms will be:

$$P_2 - E_{L,1} [P_2] = \beta_{v_2,2} v_2,$$

$$X - E_{L,1} [X] = X - E_{l,2} [X] = \varepsilon.$$

The optimal position during the final trading period

$$\begin{aligned}
q_{L,2} &= \frac{E_{L,1} [X - P_2] + c q_{L,1}}{a \sigma_\varepsilon^2 + c} - \frac{\beta_{v_2} v_2}{a \sigma_\varepsilon^2 + c} \\
&= E_{L,1} [q_{L,2}] - \frac{\beta_{v_2}}{a \sigma_\varepsilon^2 + c} v_2.
\end{aligned}$$

From this, we can calculate the variance

$$\begin{aligned}
\text{Var}_{L,1} [w_L] &= \text{Var}_{L,1} \left[\begin{aligned} &q_{L,1} \beta_{v_2} v_2 + \text{E}_{L,2} [q_{L,2}] \varepsilon - \frac{\beta_{v_2}}{a\sigma_\varepsilon^2 + c} v_2 \varepsilon \\ &+ \left(\text{E}_{L,1} [q_{L,2}] - \frac{\beta_{v_2}}{a\sigma_\varepsilon^2 + c} v_2 \right)^2 (a\sigma_\varepsilon^2 + \frac{c}{2}) \end{aligned} \right] \\
&= \text{Var}_{L,1} \left[\begin{aligned} &\left(\frac{c}{a\sigma_\varepsilon^2 + c} q_{L,1} + \frac{\text{E}_{L,1} [X - P_2]}{a\sigma_\varepsilon^2 + c} \right) \varepsilon \\ &+ \left(q_{L,1} \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 - 2\text{E}_{L,1} [X - P_2] \frac{a\sigma_\varepsilon^2 + \frac{c}{2}}{(a\sigma_\varepsilon^2 + c)^2} \right) \beta_{v_2} v_2 \\ &- \frac{\beta_{v_2}}{a\sigma_\varepsilon^2 + c} v_2 \varepsilon + \frac{a\sigma_\varepsilon^2 + \frac{c}{2}}{(a\sigma_\varepsilon^2 + c)^2} \beta_{v_2}^2 v_2^2 \end{aligned} \right]
\end{aligned}$$

and using the normality and independence of ε and v_2 ,

$$\begin{aligned}
\text{Var} [w_L] &= \text{Var} [av + b\varepsilon + cv^2 + dv\varepsilon] \\
&= a^2 \sigma_v^2 + b^2 \sigma_\varepsilon^2 + 2c^2 \sigma_v^4 + d^2 \sigma_v^2 \sigma_\varepsilon^2
\end{aligned}$$

we can write

$$\begin{aligned}
\text{Var}_{L,1} [w_L] &= \left(\frac{c}{a\sigma_\varepsilon^2 + c} q_{i,1} + \frac{\text{E}_{L,1} [X - P_2]}{a\sigma_\varepsilon^2 + c} \right)^2 \sigma_\varepsilon^2 \\
&+ \left(q_{L,1} \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 - \text{E}_{L,1} [X - P_2] \frac{2a\sigma_\varepsilon^2 + c}{(a\sigma_\varepsilon^2 + c)^2} \right)^2 \beta_{v_2}^2 \sigma_v^2 \\
&+ \left\{ \frac{\beta_{v_2}}{a\sigma_\varepsilon^2 + c} \right\}^2 \sigma_v^2 \sigma_\varepsilon^2 + 2 \left\{ \frac{a\sigma_\varepsilon^2 + \frac{c}{2}}{(a\sigma_\varepsilon^2 + c)^2} \right\}^2 \beta_{v_2,2}^4 \sigma_v^4.
\end{aligned}$$

To calculate the demand function, we need to evaluate the first derivative

$$\begin{aligned} \frac{\partial}{\partial q_{1,l}} \text{Var}_{i,1} [w_i] &= 2 \frac{c}{a\sigma_\varepsilon^2 + c} \left(\frac{c}{a\sigma_\varepsilon^2 + c} q_{i,1} + \frac{\text{E}_{L,1} [X - P_2]}{a\sigma_\varepsilon^2 + c} \right) \sigma_\varepsilon^2 \\ &\quad + 2 \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \left(q_{i,1} \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right) \beta_{v_2}^2 \sigma_v^2 \\ &\quad - 2 \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \left(\text{E}_{L,1} [X - P_2] \frac{2a\sigma_\varepsilon^2 + c}{(a\sigma_\varepsilon^2 + c)^2} \right) \beta_{v_2}^2 \sigma_v^2 \end{aligned}$$

and calculate the term

$$\begin{aligned} \frac{a}{2} \frac{\partial \text{Var}_{L,1} [w_L]}{\partial q_{1,l}} &= a \left(\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^4 \beta_{v_2}^2 \sigma_v^2 + \left(\frac{c}{a\sigma_\varepsilon^2 + c} \right)^2 \sigma_\varepsilon^2 \right) q_{i,1} \\ &\quad - a \left(\frac{(2a\sigma_\varepsilon^2 + c) a^2 \sigma_\varepsilon^4 \beta_{v_2}^2 \sigma_v^2}{(a\sigma_\varepsilon^2 + c)^4} - \frac{c\sigma_\varepsilon^2}{(a\sigma_\varepsilon^2 + c)^2} \right) \text{E}_{L,2} [X - P_2] \end{aligned}$$

The optimal portfolio for the long-term speculator is then

$$q_{L,1}^* = \frac{\text{E}_{L,1} [P_2 - P_1] + \left\{ \frac{c}{a\sigma_\varepsilon^2 + c} + a \frac{(2a\sigma_\varepsilon^2 + c) \beta_{v_2,2}^2 \sigma_v^2 a^2 \sigma_\varepsilon^4}{(a\sigma_\varepsilon^2 + c)^4} \right\} \text{E}_{L,1} [X - P_2] + cq_o}{a \left\{ \left(\frac{c}{a\sigma_\varepsilon^2 + c} \right)^2 \sigma_\varepsilon^2 + \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^4 \beta_{v_2,2}^2 \sigma_v^2 \right\} + c \left(1 - \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)}$$

which can be written as in equation (3.17)

$$\begin{aligned} q_{L,1}^* &= \frac{\text{E}_{L,1} [P_2 - P_1] + \Gamma_l \text{E}_{L,1} [X - P_2] + cq_o}{a\Omega + c \left(1 - \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)} \\ &= \frac{(1 - \Gamma) \text{E}_{L,1} [P_2 - P_1] + \Gamma \text{E}_{L,1} [X - P_1] + cq_o}{a\Omega + c \left(1 - \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)} \end{aligned}$$

The variance term, Ω is a linear combination of the uncertainty in next period's price (σ_v^2) and uncertainty in the final payout (σ_ε^2)

$$\Omega = \underbrace{\left(\frac{c}{a\sigma_\varepsilon^2 + c} \right)^2 \sigma_\varepsilon^2}_{\text{variance of } X} + \underbrace{\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^4 \beta_{v_2,2}^2 \sigma_v^2}_{\text{variance in } P_2}.$$

The sensitivity to next period's expected return is

$$\Gamma = \underbrace{\frac{c}{a\sigma_\varepsilon^2 + c}}_{\text{return next period}} + a \underbrace{\frac{(2a\sigma_\varepsilon^2 + c) a^2 \sigma_\varepsilon^4}{(a\sigma_\varepsilon^2 + c)^4} \beta_{v_2,2}^2 \sigma_v^2}_{\text{prefer to avoid uncertain } v_2}.$$

The weight Γ that the investor tilts toward the long-horizon return will always be positive, and its magnitude will increase with transaction costs. The relationship with transaction costs comes from the investor recognizing positions taken now will persist later. Additionally, there is some uncertainty in the price next period, so investors have an incentive to lock in P_1 now rather than pay an uncertain P_2 .

SHORT-HORIZON INVESTORS IN PERIOD 1

For the short-run investors, the uncertain terms will be

$$P_2 - E_{N,1}[P_2] = \beta_4 e_S + \beta_{v_2} v_2$$

and

$$X - E_{s,1}[X] = e_S + \varepsilon,$$

where

$$e_S = (\theta_2 - E_{S,1}[\theta_2]).$$

The optimal portfolio in the final trading period can then be expressed as

$$\begin{aligned} q_{S,2} &= \frac{E_{S,2}[X - P_2] + cq_{S,1}}{a\sigma_\varepsilon^2 + c} \\ &= \frac{E_{S,1}[X - P_2] + cq_{S,1}}{a\sigma_\varepsilon^2 + c} + \frac{E_{S,2}[X - P_2] - E_{S,1}[X - P_2]}{a\sigma_\varepsilon^2 + c} \\ &= E_{S,1}[q_{S,2}] + \frac{e_S(1 - \beta_4)}{a\sigma_\varepsilon^2 + c} - \frac{\beta_{v_2}v_2}{a\sigma_\varepsilon^2 + c}. \end{aligned}$$

So we can calculate the variance as

$$\text{Var}_{S,1}[w_S] = \text{Var}_{S,1}\left[q_{S,1}P_2 + q_{S,2}(X - E_{S,2}[X]) + q_{S,2}^2\left(a\sigma_\varepsilon^2 + \frac{c}{2}\right)\right]$$

So the variance is

$$\begin{aligned} \text{Var}_{S,1}[w_S] &= \left(q_{S,1} - 2\frac{(1 - \beta_4)E_{S,1}[q_{S,2}](a\sigma_\varepsilon^2 + \frac{c}{2})}{a\sigma_\varepsilon^2 + c}\right)^2 \beta_4^2 \sigma_{S,1}^2 \\ &\quad + \left(q_{S,1} - 2\frac{E_{S,1}[q_{S,2}](a\sigma_\varepsilon^2 + \frac{c}{2})}{a\sigma_\varepsilon^2 + c}\right)^2 \beta_{v_2}^2 \sigma_v^2 \\ &\quad + (E_{S,1}[q_{S,2}])^2 \sigma_\varepsilon^2 \\ &\quad + \text{Var}_{S,1}\left[\left(\frac{e_S(1 - \beta_4)}{a\sigma_\varepsilon^2 + c} - \frac{\beta_{v_2}v_2}{a\sigma_\varepsilon^2 + c}\right)^2 (a\sigma_\varepsilon^2 + \frac{c}{2})\right. \\ &\quad \left.+ \left(\frac{e_S(1 - \beta_4)}{a\sigma_\varepsilon^2 + c} - \frac{\beta_{v_2}v_2}{a\sigma_\varepsilon^2 + c}\right) \varepsilon\right] \end{aligned}$$

and we can substitute in $E_{S,1} [q_{S,2}] = \frac{E_{S,1} [X - P_2] + cq_{S,1}}{a\sigma_\varepsilon^2 + c}$ and get first derivative

$$\begin{aligned} \frac{\partial \text{Var}_{S,1} [w_S]}{\partial q_{S,1}} &= 2q_{S,1} \left\{ \left(\frac{a^2 \sigma_\varepsilon^4}{(a\sigma_\varepsilon^2 + c)^2} \right)^2 \beta_{v_2}^2 \sigma_v^2 + \left(\frac{c}{a\sigma_\varepsilon^2 + c} \right)^2 \sigma_\varepsilon^2 \right\} \\ &+ 2q_{S,1} \left\{ \left(\frac{a^2 \sigma_\varepsilon^4 - \beta_4 (2ac\sigma_\varepsilon^2 + c^2)}{(a\sigma_\varepsilon^2 + c)^2} \right)^2 \beta_4^2 \sigma_{S,1}^2 \right\} \\ &- 2E_{S,1} [X - P_2] \left(\frac{\frac{a^2 \sigma_\varepsilon^4 (2a\sigma_\varepsilon^2 + c)}{(a\sigma_\varepsilon^2 + c)^4} \beta_{v_2}^2 \sigma_v^2 - \frac{c\sigma_\varepsilon^2}{(a\sigma_\varepsilon^2 + c)^2}}{+ \frac{(a^2 \sigma_\varepsilon^4 - \beta_4 (2ac\sigma_\varepsilon^2 + c^2)) (1 - \beta_4) (2a\sigma_\varepsilon^2 + c)}{(a\sigma_\varepsilon^2 + c)^4} \beta_4^2 \sigma_{S,1}^2} \right) \end{aligned}$$

The optimal portfolio can be expressed in a form analogous to the long-run demand function in equation (3.17) by naming the short-horizon parameters, Γ_s and Ω_s ,

$$q_{S,1}^* = \frac{(1 - \Gamma_s) E_{S,1} [P_2 - P_1] + \Gamma_s E_{S,1} [X - P_1] + cq_o}{a\Omega_s + c \left(1 + \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)}.$$

The intuition and form are nearly identical, with the short-horizon investors tilting slightly more toward the long-run return, $E_{S,1} [X - P_2]$, due to their uncertainty about θ_2 ,

$$\Gamma_s = \Gamma + a \underbrace{\frac{(1 - \beta_4) (a^2 \sigma_\varepsilon^4 - \beta_4 (2ac\sigma_\varepsilon^2 + c^2))}{(a\sigma_\varepsilon^2 + c)^4} \beta_4^2 \sigma_{e_1}^2}_{\text{prefer to avoid uncertain } e_1}.$$

Their associated uncertainty term, Ω_s , is

$$\Omega_s = \left(\frac{c}{a\sigma_\varepsilon^2 + c} \right)^2 \sigma_\varepsilon^2 + \frac{(a^2 \sigma_\varepsilon^4 - \beta_4 (2ac\sigma_\varepsilon^2 + c^2))^2 \beta_4^2 \sigma_{S,1}^2 + a^4 \sigma_\varepsilon^8 \beta_{v_2}^2 \sigma_v^2}{(a\sigma_\varepsilon^2 + c)^4}.$$

UNINFORMED INVESTORS IN PERIOD 1

The uninformed investors have the highest degree of uncertainty. In period 1, this is summarized by the uncertain terms:

$$X - E_{N,2} [X] = e_1 + e_2 + \varepsilon$$

where the errors in expectations in the final period are expressed as

$$\begin{aligned} e_1 &= (\theta_1 - E_{N,2} [\theta_1]) \\ e_2 &= (\theta_2 - E_{N,2} [\theta_2]) . \end{aligned}$$

The additional, orthogonal error in the first period expectation is

$$\begin{aligned} \Delta e_1 &= (\theta_1 - E_{N,1} [\theta_1]) - (\theta_1 - E_{N,2} [\theta_1]) \\ \Delta e_2 &= (\theta_2 - E_{N,1} [\theta_2]) - (\theta_2 - E_{N,2} [\theta_2]) \end{aligned}$$

so that

$$P_2 - E_{N,1} [P_2] = \beta_3 (e_1 + \Delta e_1) + \beta_4 (e_2 + \Delta e_2) + \beta_{v_2} v_2$$

and

$$X - E_{N,1} [X] = (e_1 + \Delta e_1) + (e_2 + \Delta e_2) + \varepsilon.$$

The optimal portfolio in the final trading period can then be expressed as

$$\begin{aligned}
q_{N,2} &= \frac{E_{N,2} [X - P_2] + cq_{N,1}}{a\text{Var}_{N,2} [X] + c} \\
&= \frac{E_{N,1} [X - P_2] + cq_{N,1}}{a\text{Var}_{N,2} [X] + c} + \frac{E_{N,2} [X - P_2] - E_{N,1} [X - P_2]}{a\text{Var}_{N,2} [X] + c} \\
&= \frac{E_{N,1} [X - P_2] + cq_{N,1}}{a\text{Var}_{N,2} [X] + c} + \frac{E_{N,2} [X] - E_{N,1} [X] - P_2 - E_{N,1} [P_2]}{a\text{Var}_{N,2} [X] + c} \\
&= E_{N,1} [q_{N,2}] + \frac{\Delta e_1 + \Delta e_2 - \beta_3 (e_1 + \Delta e_1) - \beta_4 (e_2 + \Delta e_2) - \beta_{v_2} v_2}{a\text{Var}_{N,2} [X] + c} \\
&= E_{N,1} [q_{N,2}] + \frac{\Delta e_1 (1 - \beta_3) + \Delta e_2 (1 - \beta_4) - \beta_3 e_1 - \beta_4 e_2 - \beta_{v_2} v_2}{a\text{Var}_{N,2} [X] + c}.
\end{aligned}$$

The uncertainty from the perspective of the investors who acquire no information will be

$$\text{Var}_{N,1} [w_N] = \text{Var}_{N,1} \left[q_{N,1} P_2 + q_{N,2} (X - E_{N,2} [X]) + q_{N,2}^2 \left(a\text{Var}_{N,2} [X] + \frac{c}{2} \right) \right].$$

Splitting out the terms,

$$\begin{aligned}
\text{Var}_{N,1} [w_N] = & \left\{ q_{N,1} \beta_3 + E_{N,1} [q_{N,2}] - 2\beta_3 E_{N,1} [q_{N,2}] \frac{a \text{Var}_{N,2} [X] + \frac{c}{2}}{a \text{Var}_{N,2} [X] + c} \right\}^2 \sigma_{e_1}^2 \\
& + \left\{ q_{N,1} \beta_4 + E_{N,1} [q_{N,2}] - 2\beta_4 E_{N,1} [q_{N,2}] \frac{a \text{Var}_{N,2} [X] + \frac{c}{2}}{a \text{Var}_{N,2} [X] + c} \right\}^2 \sigma_{e_2}^2 \\
& + \left\{ q_{N,1} \beta_3 + (1 - \beta_3) 2E_{N,1} [q_{N,2}] \frac{a \text{Var}_{N,2} [X] + \frac{c}{2}}{a \text{Var}_{N,2} [X] + c} \right\}^2 \sigma_{\Delta e_1}^2 \\
& + \left\{ q_{N,1} \beta_4 + (1 - \beta_4) 2E_{N,1} [q_{N,2}] \frac{a \text{Var}_{N,2} [X] + \frac{c}{2}}{a \text{Var}_{N,2} [X] + c} \right\}^2 \sigma_{\Delta e_2}^2 \\
& + \left\{ q_{N,1} - 2E_{N,1} [q_{N,2}] \frac{a \text{Var}_{N,2} [X] + \frac{c}{2}}{a \text{Var}_{N,2} [X] + c} \right\}^2 \beta_{v_2}^2 \sigma_{v_2}^2 \\
& + \{ E_{N,1} [q_{N,2}] \}^2 \sigma_\varepsilon^2 \\
& + \{ \text{the terms without } q_{N,1} \} ,
\end{aligned}$$

and taking the first derivative yields the comon form

$$q_{N,1}^* = \frac{(1 - \Gamma_N) E_{N,1} [P_2 - P_1] + \Gamma_N E_{N,1} [X - P_1] + c q_0}{a \Omega_N + c \left(1 + \left(\frac{a \sigma_\varepsilon^2}{a \sigma_\varepsilon^2 + c} \right)^2 \right)}.$$

In this case,

$$\begin{aligned}
\Omega_N = & \left(\frac{c}{a \text{Var}_{N,2} [X] + c} + \beta_3 \left(\frac{a \text{Var}_{N,2} [X]}{a \text{Var}_{N,2} [X] + c} \right)^2 \right)^2 \sigma_{e_1}^2 \\
& + \left(\frac{c}{a \text{Var}_{N,2} [X] + c} + \beta_4 \left(\frac{a \text{Var}_{N,2} [X]}{a \text{Var}_{N,2} [X] + c} \right)^2 \right)^2 \sigma_{e_2}^2 \\
& + \left(1 - \beta_3 \left(\frac{a \text{Var}_{N,2} [X]}{a \text{Var}_{N,2} [X] + c} \right)^2 \right)^2 \sigma_{\Delta e_1}^2 + \left(1 - \beta_4 \left(\frac{a \text{Var}_{N,2} [X]}{a \text{Var}_{N,2} [X] + c} \right)^2 \right)^2 \sigma_{\Delta e_2}^2 \\
& + \left(1 - \left(\frac{a \text{Var}_{N,2} [X]}{a \text{Var}_{N,2} [X] + c} \right)^2 \right)^2 \beta_{v_2}^2 \sigma_{v_2}^2 + \left\{ \frac{c}{a \text{Var}_{N,2} [X] + c} \right\}^2 \sigma_\varepsilon^2
\end{aligned}$$

and

$$\begin{aligned}
\Gamma_N = & \frac{c}{a\text{Var}_{N,2}[X] + c} + c \frac{a\text{Var}_{N,2}[X]}{(a\text{Var}_{N,2}[X] + c)^2} \\
& + a \left\{ \frac{\begin{aligned} & (c(a\text{Var}_{N,2}[X] + c) + \beta_3(a\text{Var}_{N,2}[X])^2) \\ & \times (a\text{Var}_{N,2}[X] + c - \beta_3(2a\text{Var}_{N,2}[X] + c)) \end{aligned}}{(a\text{Var}_{N,2}[X] + c)^4} \right\} \sigma_{e_1}^2 \\
& + a \left\{ \frac{\begin{aligned} & (c(a\text{Var}_{N,2}[X] + c) + \beta_4(a\text{Var}_{N,2}[X])^2) \\ & \times (a\text{Var}_{N,2}[X] + c - \beta_4(2a\text{Var}_{N,2}[X] + c)) \end{aligned}}{(a\text{Var}_{N,2}[X] + c)^4} \right\} \sigma_{e_2}^2 \\
& + a \left\{ \begin{aligned} & \left(1 - \beta_3 \left(\frac{a\text{Var}_{N,2}[X]}{a\text{Var}_{N,2}[X] + c} \right)^2 \right) \\ & \times \left(\frac{(1 - \beta_3)(2a\text{Var}_{N,2}[X] + c)}{(a\text{Var}_{N,2}[X] + c)^2} \right) \end{aligned} \right\} \sigma_{\Delta e_1}^2 \\
& + a \left\{ \begin{aligned} & \left(1 - \beta_4 \left(\frac{a\text{Var}_{N,2}[X]}{a\text{Var}_{N,2}[X] + c} \right)^2 \right) \\ & \times \left(\frac{(1 - \beta_4)(2a\text{Var}_{N,2}[X] + c)}{(a\text{Var}_{N,2}[X] + c)^2} \right) \end{aligned} \right\} \sigma_{\Delta e_2}^2 \\
& - a \left\{ \left(1 - \left(\frac{a\text{Var}_{N,2}[X]}{a\text{Var}_{N,2}[X] + c} \right)^2 \right) \left(\frac{2a\text{Var}_{N,2}[X] + c}{(a\text{Var}_{N,2}[X] + c)^2} \right) \right\} \beta_{v_2}^2 \sigma_{v_2}^2 \\
& + a \left\{ \frac{c}{(a\text{Var}_{N,2}[X] + c)^2} \right\}^2 \sigma_{\varepsilon}^2
\end{aligned}$$

MARKET CLEARING AND INVESTOR EXPECTATIONS

The investors will form expectations about investment prospects (X) and the effect of the noise shocks (v_1 and v_2) from the market price. Intuitively, investor expectations

of θ increase in the market price, but larger noise shocks dampens this relationship. It remains to be verified that the assumed linear relationship between prices and the unknown variables as suggested in equations (3.8) and (3.9) holds.

In period 1, the market clears when

$$Q_o = \lambda_N q_{N,1} + \lambda_S q_{S,1} + \lambda_L q_{L,1} + \frac{v_1}{a\sigma_\varepsilon^2 + c}.$$

The demand functions for the short-horizon and long-horizon investors are both linear in $E[X]$ and hence linear in the state variables, so substituting them into the market clearing condition shows the price to be linear in the state variables. The expectations of the risky payout will all be linear in P_1 , which can be seen from substituting in the demand functions to the market clearing condition

$$\begin{aligned} P_1 \propto & \left\{ \lambda_S \left(\frac{(1 - \Gamma_S) \beta_3 + \Gamma_S}{a\Omega_S + c \left(1 + \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)} \right) + \lambda_L \left(\frac{(1 - \Gamma) \beta_3 \theta_1 + \Gamma \theta_1}{a\Omega + c \left(1 - \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)} \right) \right\} \theta_1 \\ & \left\{ \lambda_L \left(\frac{(1 - \Gamma) \beta_4 + \Gamma}{a\Omega + c \left(1 - \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c} \right)^2 \right)} \right) \right\} \theta_2 \\ & + \left\{ \frac{1}{a\sigma_\varepsilon^2 + c} \right\} v_1 \end{aligned}$$

This confirms (3.8).

Similarly, in period 2 the market clearing condition shows that

$$P_2 \propto \frac{\lambda_S + \lambda_L}{a\sigma_\varepsilon^2 + c} (\theta_1 + \theta_2) + \left\{ \frac{1}{a\sigma_\varepsilon^2 + c} \right\} v_2,$$

which confirms (3.9).

THE IMPACT OF MORE EFFICIENT TRANSACTIONS

Let's now turn to the question of what happens if the financial sector is more operationally efficient and the cost of transacting decreases. I consider two key comparative statics: how does this affect total active investment management ($\frac{\partial \sum \lambda_i}{\partial c}$) and how does this effect differ by investment horizon ($\frac{\partial \lambda_s}{\partial c}$ versus $\frac{\partial \lambda_l}{\partial c}$).

Proposition 2 (More active management) *As the cost of transacting decreases, total informed trading increases,*

$$\frac{\partial \sum \lambda_i}{\partial c} \leq 0$$

and this becomes a strict inequality if there is any interior solution (i.e. $0 < \lambda_j < 1$ for some j).

The value gained from information lies in the ability capitalize on the information through active trading. Clearly, in the limiting case, $\lim_{c \rightarrow \infty} \lambda_n^* \rightarrow 1$. For interior solutions, we must consider the marginal impact of transaction costs on the relative utility of informed and uninformed investors. The unconditional expected utility of an informed speculator will be a decreasing, continuous function of transaction costs. The unconditional expected utility of a passive investors will also decrease—but much less rapidly. Hence, $\frac{\partial \sum \lambda_N}{\partial c} \geq 0$. Since these functions are continuous, equality will only hold in the corner solutions where marginal changes in expected utility have no effect on the allocations of investor type.

Proposition 3 (Shorter investment horizons) *Lower transaction costs have a greater effect on short-horizon investors than long-horizon investors,*

$$\frac{\partial \lambda_S}{\partial c} \leq \frac{\partial \lambda_L}{\partial c}$$

with strict inequality for interior solutions (i.e. $\lambda_L \in (0, 1)$ and $\lambda_S \in (0, 1)$).

This result comes from the fact that the short-horizon investors' optimal portfolio contains a subset of the information of the long-horizon investor. So the desire to spread trading over a longer horizon is offset by the fact that the short-horizon signal in period 1 (θ_1) may be in the opposite direction as the signal in period 2 (θ_2). As a result, short-horizon traders are forced to trade more for the same expected return.

In fact, in a model with many periods (T large), the short-horizon traders will find that the independence of θ_t makes trading in the earliest periods costly relative to the weakness of their accumulated signal. As the final horizon approaches, the short-horizon traders will be more inclined to trade as their accumulated signal is stronger and less likely to suggest they need to unwind their trades because of future information.

In contrast, the long-horizon traders are eager to trade on their information as early as possible, but they submit to spreading their trading across later periods in their desire to minimize their transaction costs. There are also information advantages to spreading out trades, since larger trades move prices and allow other traders to freely infer the costly information, but the infinitesimal traders do not absorb this externality.

3.2 EXPLAINING THE EMPIRICAL GROWTH IN CAPITAL MARKET SPENDING

A key contribution of this paper is document the relationship between the efficiency of financial transactions and the growth of modern finance. As improvements in technology and market organization make transactions less costly, we should expect to see the volume of transactions increase. This simply follows from the economic Law of Demand. A more surprising result is that as financial costs decrease, total spending on finance increases. This is fundamentally a statement about elasticities.

In this section, I focus on establishing the relationship between financial efficiency and the aggregate measures of financial spending and activity. I use timing to assert causality in the Granger sense, and using the (plausibly) exogenous historical break in May of 1975. The evidence is statistically strong but open to the criticism that the changes in efficiency may be interrelated with contemporaneous events. In section 3.3, I will use cross-sectional variation in the panel data to establish even stronger results and focus more explicitly on measuring the information content and investment horizon, two key features of the model.

A TIME SERIES OF TRANSACTION COSTS

With the possible exception of the very recent past, brokerage commissions were the primary cost in trading equities (Berkowitz, Logue and Noser, 1988). They funded all the operations required in financial market transactions. To test the efficiency explanation for the growth of capital markets, I construct a historical time series that

measures the representative cost of transacting. The measure I propose splices two date ranges: 1927-1975 and 1975-2010.

PRE-1975: THE NYSE FIXED COMMISSION SCHEDULE

From its founding in 1792 up to 1975, the New York Stock Exchange (NYSE) enforced a minimum commission schedule on all of its member firms. The smaller, regional exchanges mirrored the commission schedule of the NYSE, and in the rare cases where they didn't, they faced enormous industry pressure to conform. The stated goal was to "prevent competition amongst the members" to protect their profits. Exchange members referenced the general fear of unfettered trading and defended high trading costs by observing that "a very low or competitive rate would also promote speculation."⁴

An example commission schedule, corresponding to the NYSE rates for 1956 is displayed in Figure 3.4. We can see how the formula defining the commission rate is a function of the nominal share price. Purchasing a round lot (100 shares) of a stock costing \$30 per share, for example, would have a commission of \$15 + 0.5 times \$30. A round lot of a \$60 stock would cost \$35 + 0.1 times \$60.

To construct a time series of the average transaction cost prior to 1975 I collect the NYSE commission schedules, including the NYSE annual fact books and the monthly S&P Stock Owners Guide. Combining these commission schedules with trading volume and price data from CRSP,⁵ I construct an annual series of the

⁴Report of the Committee Appointed Pursuant to House Resolutions 429 and 504 to Investigate the Concentration of Control of Money and Credit, H.R. REP. NO. 62-1593

⁵Center for Research in Security Prices. Graduate School of Business, The University of Chicago

MINIMUM COMMISSION RATES		EXAMPLES of Minimum Commission Charges for 100 Shares
Price per Share	Minimum Commission Rate per 100 Shares	
Between \$1 and \$20	\$5 plus price per share	At \$15 a share, the commission would be \$5 plus \$15 or \$20.
From \$20 to \$50	\$15 plus 1/2 price per share	At \$35 a share, the commission would be \$15 plus \$17.50 or \$32.50.
Between \$50 and \$150	\$35 plus 1/10 price per share	At \$100 per share, the commission would be \$35 plus \$10 or \$45.
At \$150 or more	\$50	At \$200 per share, the commission would be \$50.

Figure 3.4: NYSE Commission Schedule, 1956

An image of the New York Stock Exchange minimum commission schedule for 1956, as reported on page 7 of the NYSE Fact Book for 1965.

weighted average cost of trading.

MAY DAY 1975

In the aftermath of the financial disasters surrounding the Great Depression, the Securities Exchange Act of 1934 charged the Securities and Exchange Commission (SEC) with regulating and approving changes to any enforced commission schedules. Over the following forty years, the NYSE would periodically submit proposals to increase rates. A pattern emerged whereby the NYSE would complain about the rising costs and shrinking profits of its members, propose an increase in the commission schedule in order to maintain an appropriate level of profitability, and they would get immediate approval from the SEC.

In 1968, however the SEC scrutinized the latest proposed increase with more skepticism. Regulators asked why the cost of transacting in the financial markets could not itself be the product of a competitive response. The response from the exchange was emphatic: "One does not move the keystone of an industry which facilitates the raising of the bulk of new capital for this country...Negotiated rates would bring a degree of destructive competition."⁶

Although the SEC continued to approve a series of regular increases, this initial dissatisfaction was not placated. On January 23, 1975 the SEC adopted rule 19-b, requiring all stock exchanges to end the practice of the fixed commission schedule and allow members to set rates competitively. This rule was to go in effect on May 1, 1975. Distressed brokers and the popular press referred to the deadline as May Day.

(2012), Used with permission. All rights reserved.

⁶Richard Hack, NYSE president (August 19, 1968)

As brokers competed for the first time on trading costs, there was a sharp drop in costs, especially for institutional investors. The SEC instituted a number of studies trying to measure the impact of their rule. Only two weeks after the beginning of competitive rates, the SEC Commissioner noted that they “have seen sharp price cutting, in some instances to half or less of previously prevailing rates.”⁷ The SEC study of 1978 concluded that institutional trading costs had stabilized to a level 52.9% below their fixed rate levels.⁸ Interestingly, the costs to individual traders decreased only moderately, giving rise to price discrimination among investor types (Tinic and West, 1980).

POST-1975: NYSE MEMBER FINANCIAL STATEMENTS

To continue the time series measuring the cost of transacting in the modern period of negotiated commissions post-1975, I collect commission revenues from the member financial statements of the NYSE and divide them by trading volume to estimate the weighted average cost per share.

Figure 3.5 shows the composite time series from 1927 to 2010. We can see the significant increase in the early 1930s followed by a relatively steady increase in costs for almost 50 years until the sudden drop resulting from the events of May 1975. To ensure the aggregate time series is a fair representation of aggregate transaction costs, I compare it to a number of independent measures. These include: the survey results from Greenwich Associates, a consultancy that surveys institutional investors

⁷Remarks by A. A. Sommer Jr. in a talk titled “The New Breath of Competition” delivered at the Seminar on the Analysis of Security Prices, University of Chicago, May 15, 1975.

⁸SEC Staff Report on the Securities Industry in 1978

regarding the costs they pay for their transactions; the SEC studies measuring transaction costs in the wake of rule 19-b; and for historical purposes, the cost associated with trading a \$30 stock, holding the nominal share price constant through the duration of the fixed commission schedule. Each of these measures corresponds relatively closely to the composite series I created.

Since the post-1975 series imputes costs rather than calculating them directly, it is especially useful to compare them with data published by Greenwich Associates, a firm that has been polling institutional investors on their average commission costs since 1976. The time series of their survey results is plotted in green triangles alongside my own estimates on Figure 3.5. The two series are highly similar, except in the first few years of the sample where the commissions paid by institutions are even lower than the computed average. This is consistent with historical reports that the trading commissions charged to individuals did not drop immediately in response to the deregulation until the advent of discount stock brokers around 1980.

Looking at the data prior to 1975, I plot the evolution of the cost of trading a \$30 stock using the orange squares. Historical patterns in share prices and trading volume cause the higher frequency variation in my composite series, making it useful to compare against a series where the nominal share price is held constant. Any changes can then be attributed to the imposed cost schedule and not to endogenous investor behavior. Focusing on the cost of trading a \$30 stock from 1928 to 1973, we see the round trip cost more than tripled, from 1.07% to 3.46% of the notional value. Including the additional 1.7% for paying the typical $\$1/4$ cost from the bid-ask spread, the total cost of buying and selling exceeded 5% in 1975. It is important to note the

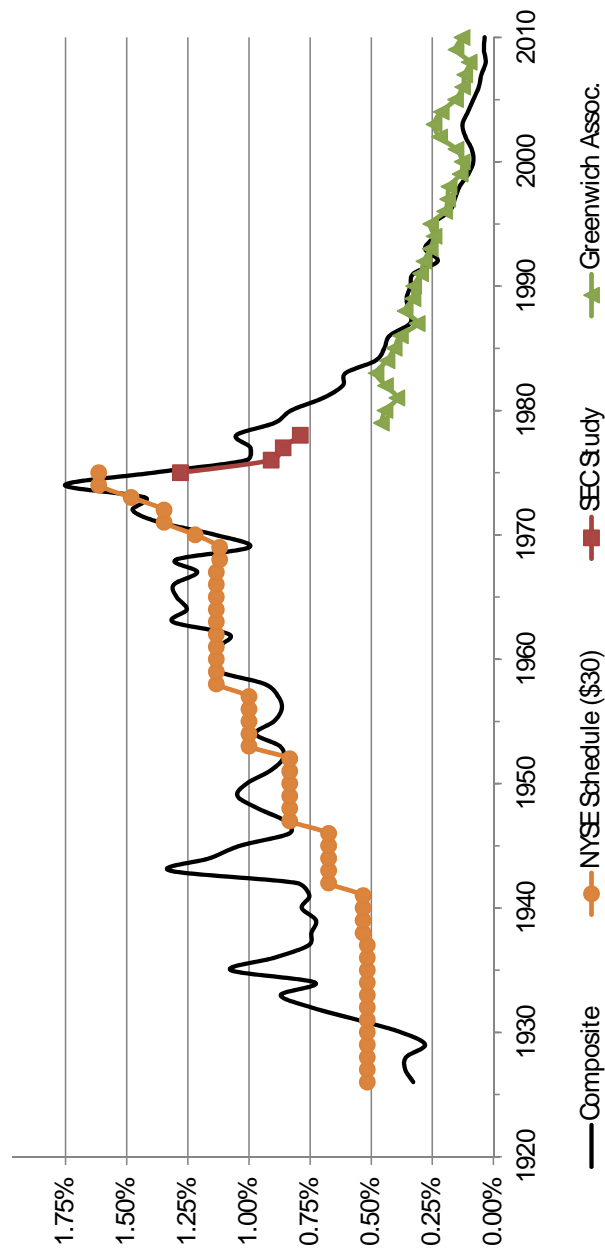


Figure 3.5: Transaction Cost Time Series

The figure above plots the composite transaction cost measure, constructed as described in section 3.2, plotted alongside three other comparison series. The orange series shows the cost of trading a stock with a nominal share price of \$30 according to the published NYSE commission schedule, the red series shows the decrease in commission costs as measured by the Securities and Exchange Commission in their analyses of the effects of commission deregulation, and the green line plots the average equity commission charge collected in a survey of institutional investors by Greenwich Associates, a financial consulting firm.

economic importance of this magnitude. To put this in perspective, the average stock response to an earnings announcement is in the range of 4%⁹, so even if it were possible to know earnings announcements with certainty, you would typically not be able to recover the cost of transacting. The costs were so high that only large misvaluations could merit attention. A speculator would favor low frequency information, with the hope that transaction costs might be amortized over a long horizon. Furthermore, any dynamic trading strategy, such as a portfolio rebalancing rule or a derivative replication, would be incredibly costly.

TIME SERIES ANALYSIS

We can expect the constructed time series of transaction costs to be negatively correlated with trading volume, a relationship that should hold true in nearly any economic model. If the proposed efficiency explanation for capital market growth plays a significant role, transaction costs should also be negatively related to capital market spending. In particular, this increase should correspond to active investment management and not just an increase in the operational costs associated with higher trading volume. Lastly, the prediction of more informed speculation also suggests that employees with higher skill and compensation enter the sector in response to a cheaper cost of transacting.

The series measuring the cost of capital markets continues to be the value added measure of capital market industries relative to private GDP with annual data from 1927 to 2010. The series measuring capital markets compensation relative to average

⁹See, for example, Francis, Schipper and Vincent (2002).

US private compensation was also previously described and plotted in Figure 3.1. I measure equity turnover by collecting all available CRSP data on stock volume and shares outstanding for common equity of US firms. Additional details behind the data sources and data construction can be found in the online data appendix.

SUMMARY STATISTICS AND SIMPLE REGRESSION ANALYSIS

The summary statistics for these four time series are presented in Table 3.2. We can see that the transaction cost, measured in basis points (hundredths of one percent), averages 71 basis points over the full sample. The series ranges significantly from more than 150 bps near its peak to just a few basis points in recent years. The fraction of GDP devoted to capital markets averages about 79 basis points over this time series, averaging about 30 basis points before 1975 and increasing to about 200 basis points in recent years. The compensation for capital market employees has an average that is approximately twice the US private sector average over the full sample, increasing to almost 4 times average compensation in recent years. Equity turnover is about 56% a year on average, suggesting an average holding period of approximately two years. While turnover was very high in the late 1920's, it was consistently low for most of the 20th century and then rises again in the recent past, with a current horizon of just a few months.

The correlations of the four series are displayed in the bottom panel of Table 3.2. As predicted, transaction costs have a strong negative relationship with the size of capital market spending and the volume of trade. While supporting the idea of a contemporaneous relationship, the slow-moving nature of all four time series might

Table 3.2: Time series summary statistics and correlations

This table shows summary statistics for annual data on: the average commission cost of transacting stocks in the United States (tcost) constructed as described in section 3.2; the percentage of national income consumed by capital markets related activity using a GDP value-added measure divided by private GDP calculated using data from the Bureau of Economic Analysis (capmkt%); the ratio of the average salary for employees in capital markets related industries relative to the average salary across all private-sector employees using data from the Bureau of Economic Analysis (comp ratio); and the annual turnover in US equities measured by dividing annual volume by shares outstanding as reported in CRSP. Annual observations are used over the period 1927-2010 to calculate the mean, standard deviation and various percentiles in the upper panel. Correlations are displayed in the lower panel.

1927-2010					
	mean	std.	1 %ile	50 %ile	99 %ile
tcost (bps)	71.1	43.6	3.6	78.4	152.0
capmkt% (bps)	78.8	65.7	8.5	43.1	221.6
comp ratio	2.09	0.77	1.20	1.72	3.92
turnover	55.7	58.9	7.3	30.4	277.1
Correlation					
	tcost	capmkt	comp	turnover	
tcost (bps)	1.00	-0.81	-0.83	-0.76	
capmkt (bps)	-0.81	1.00	0.90	0.72	
comp ratio	-0.83	0.90	1.00	0.87	
turnover	-0.76	0.72	0.87	1.00	

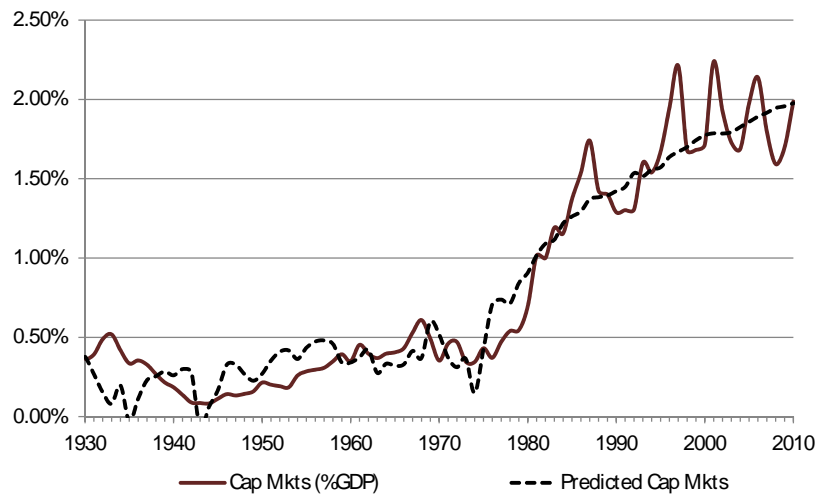


Figure 3.6: Predicting the cost of capital markets using the cost of transacting

The above figure plots in red the percentage of national income consumed by capital markets related activity using a GDP value-added measure divided by private GDP calculated using data from the Bureau of Economic Analysis. The dotted line shows the fit of a time series regression using the composite commission time series and a linear time trend.

cast doubt on the statistical significance.

We can see this more precisely in the simple regressions shown in Table 3.3, where the GDP share of capital market (*capmkt*), the relative compensation ratio for capital markets (*comp*) and the estimated US equity market turnover (*turnover*) are each regressed on the transaction cost series (*tcost*). As an illustration of the strength of this predictive relationship, Figure 3.6 plots the growth in the cost of capital markets (shown previously in Figure 3.1) against the predicted value from the regression.

While there is certainly some unexplained variation, the visual fit is striking. Note that each of these series is highly persistent, as is observed in their plots, so it comes as no surprise that an augmented Dickey-Fuller test does not reject the possibility of a

unit root. This degree of persistence would discount the significance of their observed correlations.

REGRESSION OF FIRST DIFFERENCES

To make a stronger case for this relationship and establish causality (in the Granger sense that past transaction costs forecast growth in capital market activity), we can consider how the changes in one series affects the other by taking first differences.

With the high degree of persistence in the raw time series, they may be susceptible to the type of spurious regression results that occur with unit roots. The first differences could then reveal if the time series are truly related, and if so, if one tends to forecast the other. Table 3.3 reports the results for regressions forecasting annual changes in capital market spending, the capital market compensation ratio, and trading volume as each is regressed on annual changes in transaction costs with up to 4 lags.

The predicted negative relationship remains. Interestingly, changes in transaction costs lead changes in the other series by approximately 2 to 3 years. For example, in the first regression of capital market spending on lagged changes in transaction costs we see negative coefficients for every lag with the second lag being of the strongest magnitude. We can interpret this coefficient as suggesting a one basis point decrease in the cost of transactions predicts that capital markets will consume a 13 basis point higher share of private GDP two years in the future. The same one basis point decrease in the cost of transacting would predict the average compensation of capital markets professionals in three years to rise by an additional 0.18 times the compensation of the average US employee. Looking at the effect on trading volume,

Table 3.3: Time series regressions of first differences

This table shows the results of regressing changes in the income share of capital markets (Δcapmkt), capital market compensation (Δcomp), and equity turnover by volume ($\Delta\text{turnover}$) on changes in the commission cost of stock transactions (Δtcost) with up to four lags. Newey-West adjusted t-statistics, with four lags, are reported in parentheses. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	Δcapmkt (1)	Δcomp (2)	$\Delta\text{turnover}$ (3)
Δtcost	-3.46 (4.93)	4.33 (8.45)	0.62 (6.34)
$L(\Delta\text{tcost})$	-3.01 (6.20)	3.42 (9.77)	-4.24 (5.55)
$L^2(\Delta\text{tcost})$	-12.98* (7.29)	2.62 (11.67)	-2.95 (5.64)
$L^3(\Delta\text{tcost})$	-2.41 (4.77)	-18.11** (7.07)	-9.16** (4.13)
$L^4(\Delta\text{tcost})$	-6.06 (6.56)	-7.20 (7.95)	-7.34 (5.86)
Constant	1.92 (1.17)	2.38 (1.56)	2.16 (1.72)
Observations	80	80	80

this one basis point decrease in transaction costs would suggest trading volume to be 9% higher in three years' time.

This is actually what we might predict if innovations to transaction costs are unexpected. In the context of the proposed model, investors commit to their type *ex ante*, so we would expect the delayed response to correspond to the time it takes to acquire the talent and research necessary to launch new dynamic strategies.

The statistical relationship seems compelling, although any claims about the importance of the efficiency mechanism are certainly open to critiques of omitted variable bias. A number of important regulatory and technological changes happened during the 1970's. The coincident growth in capital markets and decline in transaction costs could be coincidence, although it would be difficult to explain the strong predictive power of the transaction cost changes exhibited in Table 3.3. To strengthen the identification of the true mechanism causing financial growth, we can look at the cross-section of firms and focus on specific predictions around the events of May 1975.

3.3 MARKET ACTIVITY AND ASSET PRICES IN THE CROSS SECTION

Moving from broad statements about financial activity to the activity we observe for individual firms provides a more refined measure of how much of the growth in active investing can be explained by transaction efficiency. The model presented in section 3.1 had specific predictions regarding trading activity and the information content of asset prices. As trading efficiency increases we expect to see more trading volume and more informative asset prices. There should also be a differentially large impact on

the shorter investment horizons relative to longer horizons. Observing cross-sectional variation in the prices and trading activity of individual firms over the past few decades will generate micro-level support to add to the macro-level time series evidence presented in the previous section.

For increased confidence that we are isolating a key driving mechanism behind the growth of active investing, we can use the events of May 1975 as Rule 19-b came in force. First, we expect that the subsequent drop in transaction costs associated with competitive brokerage commissions should lead to a subsequent increase in the trading and information content of US equities. Following a key prediction of the model, we should expect this to be stronger for shorter horizons. Then, to better identify the efficiency channel, we can use specific features of how the fixed commission schedule affected the cross-section of firms until May 1975 to measure differential effects. This additional level of control helps rule out competing explanations that might have occurred on or around 1975.

CONNECTING THE PANEL DATA WITH THE STYLIZED MODEL

In the stylized model of section 3.1, the information content of long-horizon prices can be measured through the regression coefficient from projecting the risky investment outcome $(X - E[X])$ on to the change in the long-horizon price $(R_L = P_o - P_1)$, defining

$$\beta_L = \frac{\text{Cov}[X, R_L]}{\text{Var}[R_L]} = \frac{\beta_{\theta,1}\sigma_\theta^2}{\text{Var}[R_L]}.$$

Intuitively, the information content of long-horizon prices is positively related to the quantity of long-horizon active investors.¹⁰

The information content of short-horizon prices can be similarly expressed by

$$(R_S = P_1 - P_2)$$

$$\beta_S = \frac{\text{Cov}[X, R_S]}{\text{Var}[R_S]} = \frac{\beta_{\theta,2} \sigma_\theta^2}{\text{Var}[R_S]}.$$

which increases with the sum of the long-horizon and the short-horizon active investors.

We can construct an analogous measure with empirical data on stock prices and earnings. I define the "long horizon" as the period stretching from two years prior to a firm's earnings announcement to 7 months prior to the earnings announcement, the "short horizon" spanning 7 months prior to the earnings announcement to one month prior to the earnings announcement, and the "announcement period" spans from one month before to two months after the announcement. The risky investment outcome will be defined as the scaled change in a firm's quarterly earnings (Δx_t).

This motivates a corresponding empirical regression of the firm's uncertain payout on the returns over each horizon,

$$\Delta x_t = \beta_o + \beta_L \times r_L + \beta_S \times r_S + \beta_A \times r_A \quad (3.18)$$

Each of the returns will be measured as the change in log-price, so if time t is

¹⁰Formally, this can be stated as $\frac{\partial \text{Cov}[X, R_L]}{\partial \lambda_L} > 0$, and also, $\frac{\beta_L}{\partial \lambda_L} > 0$ given $\text{Var}[R_L] > \beta_{\theta,1} \sigma_\theta^2$.

measured in months relative to the earnings announcement,

$$r_L = \ln(P_{t-7}) - \ln(P_{t-24})$$

$$r_S = \ln(P_{t-1}) - \ln(P_{t-3})$$

$$r_A = \ln(P_{t+1}) - \ln(P_{t-1}).$$

Similarly, the risky payout will be measured as a log return scaled by the price observed prior to all the returns. If EPS_t corresponds to the earnings-per-share reported on the announcement date, the risky payout in the panel regressions specified by (3.18) will be defined as

$$\Delta x_t = \ln \left(1 + \frac{EPS_t - EPS_{t-3}}{P_{t-24}} \right).$$

DESCRIPTION OF PANEL DATA

For each year from 1960 to 2012, I construct a universe of firms by selecting the 1000 largest firms by market capitalization, as measured by their CRSP-reported market cap on December 31st of the prior year. For this set of firms, I collect historical weekly total returns, nominal share prices, trading volume, and shares outstanding. Using the linked CRSP-Compustat data, I collect a panel of their reported earnings per share and the date of the earnings announcement.

The announcements dates are not always available, particularly early in the sample, so I create an additional supplemental series of earnings announcement data where I

use historical announcement patterns to estimate the date when not available. This has the advantage of increasing the sample size, and the methodology for estimating historical announcement dates appears to be very accurate when checked against firms for which the actual dates are known. Since the announcement return period is defined to begin one month prior to the reported announcement, any imprecision should have little effect on the results of the subsequent panel regressions.

Table 3.4 reports the summary statistics for the variables considered in the panel data regression. The earnings news measure (Δx_t) for these large firms over the 45 year sample averages approximately zero with a standard deviation of approximately 2%. The market price for the firms in the sample appears surprisingly high, at about \$104, but this is actually an artifact of Berkshire-Hathaway's inordinately large nominal share price. The median share price is \$32 with a standard deviation of \$24. Dividing the trading volume recorded in CRSP for each quarter by the shares outstanding, I obtain firm-level annualized turnover rates for each firm-quarter in the panel. Over the full sample, annualized turnover averages 2.36, with a wide degree of variation across firms. The return variables, r_L , r_S and r_A , each correspond to a different horizon length, so the magnitudes of their average returns and standard deviations are not directly comparable.

The lower panel of Table 3.4 reports the same summary statistics for the sub-sample corresponding to the five years before May of 1975, the two years of observations that overlap with May 1975, and five years afterward. This subsample, and ones like it, will be used in the panel regressions where the data window tightens around the events around the implementation of Rule 19-b.

Table 3.4: Summary Statistics for Panel Data Analysis

The summary statistics below are for the quarterly panel data collected for the 1,000 firms in the annual universe being analyzed. The universe is reset each year, taking the 1,000 largest firms by market cap. The first panel cover the full sample period, while the lower panel covers the 5-year window before fixed exchange regime was ended on May 1, 1975 up until 5-years after May 1, 1977—the date at which none of the collected series overlap with the fixed-rate commission regime.

	mean	std.	1 %ile	50 %ile	99 %ile
1966 - 2010					
Δx_t	-0.01	2.08	-9.88	0.05	8.83
price	104.10	23.56	6.24	32.50	132.60
turnover	2.36	3.20	0.10	1.39	14.58
r_L	0.124	0.446	-0.835	0.082	1.526
r_S	0.013	0.216	-0.577	0.011	0.598
r_A	0.006	0.156	-0.411	0.004	0.434
(N = 134,128)					
1970-1982					
Δx_t	0.02	1.97	-7.60	0.07	6.90
price	32.17	22.26	6.75	27.38	111.80
turnover	0.81	0.87	0.05	0.58	4.19
r_L	0.102	0.388	-0.785	0.075	1.245
r_S	0.017	0.193	-0.485	0.012	0.546
r_A	0.007	0.139	-0.336	0.004	0.394
(N = 36,174)					

ROLLING PANEL REGRESSION

To generate a graphical measure of the changing information content of prices over time, we can perform a rolling panel regression. I hold the window length constant at two years and then estimate the panel regression corresponding to equation (3.18) with firm fixed effects. Figure 3.7 displays the rolling coefficient estimates as a scatterplot in the upper axis, where each estimated long horizon coefficient, β_L , corresponds to a white circle and each estimated short-horizon coefficient, β_S , correspond to a shaded circle. The lower axis reports the estimated root mean square error (RMSE) and the R-squared coefficient of each regression.

The rising pattern in the information content of asset prices is clearly visible. While the magnitude of these betas are roughly similar in the first 10 years of the sample, the predictive power of the short-horizon prices increases much more rapidly than the long-horizon prices. In a more careful subsequent regression estimating the trend in information content over time, I show the increase in the long horizon coefficient, while positive, to be statistically difficult to distinguish from a hypothesis of no change.

This is consistent with the results of Bai et al. (2012). They look at the information content of prices at one to three years prior to earnings releases. This is what my results would consider long-horizon information, and I find no compelling evidence that this information has improved over time.

On the other hand, asset prices less than one year prior to earnings announcements show a consistent increase in information content. Previewing my focus on the events of May 1975, this figure already gives a strong visual indication

that the strongest increases in information content correspond to this change as active investing increased dramatically.

While this rolling analysis is instructive, the underlying investment setting may not be fully comparable as the sample rolls across time. The information gathering problem may be different from one decade to the next, and there may be significant changes in the price-to-earnings relationship that would affect the magnitude of the coefficients.

With that in mind, it is interesting to look at the bottom axis of Figure 3.7 and note how both the explained variation (R^2) and the unexplained variation (RMSE) are increasing in the late 1970's and, to a lesser extent, over the full historical sample. This suggests that the raw difficulty of forecasting earnings increased, but so did the fraction of variation that prices could explain.

PANEL REGRESSION WITH TREND

To directly estimate the pattern of change in the information contained in asset prices over the full sample, I run a full panel regression, interacting the return variables with the time trend. The variable, *trend* is measured in years, and the coefficient on $r_L \times trend$ can be interpreted as the annual change in the regression coefficient measuring long-horizon information content. Corresponding interaction terms are used for the short-horizon and announcement return.

Table 3.5 reports the results of the base panel regressions suggested in equation (3.18) as well as a version with these time trend interactions. The reported standard errors are estimated using industry clustering, where I use the two digit SIC code as

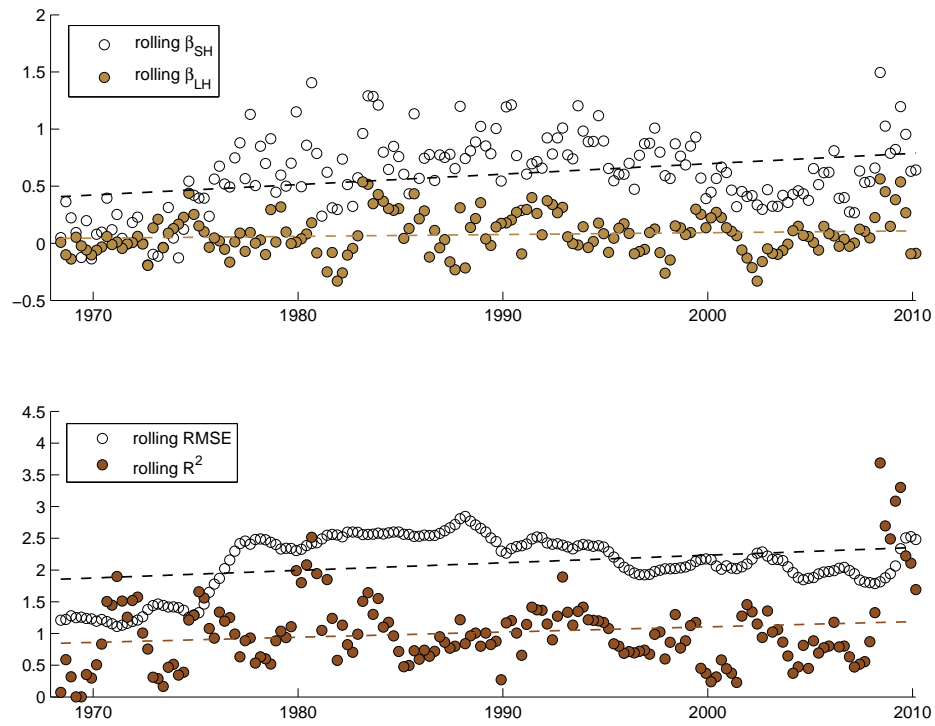


Figure 3.7: Rolling Regression Coefficient and Moving Average, 1965-2010

The two axes plot the results of the rolling regressions described in section 3.3. The top axis plots the estimated regression coefficients and the lower axis plots the square root of the mean squared error (RMSE) and the R^2 values.

the definition for industry throughout.

The regression reported in the first column of Table 3.5 reports the results of the base regression using firm fixed effects, considering variation within firms. The second regression specification uses industry and quarter fixed effects to isolate the impact of variation among similar firms in the same time period. The results of each specification are very similar. The strong statistical significance of these regression coefficients should not be too surprising; changes in asset prices correspond to present and future changes in earnings. On the other hand, the coefficient on the long-horizon return is not particularly strong in the first specification with firm fixed effects, and disappears entirely in the second specification.

The third specification is the primary one of interest. It shows the gradual change in these coefficients over time. The interaction term between the short horizon return and the time trend is statistically significant at the 1% level. In contrast the long horizon return shows little evidence of increasing informativeness over time. Of note, the three-month return around the earnings announcement actually shows a decreasing relationship in predicting the reported earnings. The fact that we observe opposite effects on the short-horizon and announcement returns may indicate a substitution of information being pulled into earlier asset prices.

THE POST-1975 EFFECT

Over such a long sample, any number of underlying parameters could be changing. The types of firms today are certainly very different than those of the 1960s. There could very well be differences in the difficulty of predicting their future profitability,

Table 3.5: Base panel regression with time trend

The regression estimates below are the result of panel regressions of earnings news (Δx defined in section 3.3 of the paper) on past log returns, log returns interacted with a time trend. The regression also includes a constant term and constant *trend* variable, but the coefficients are not reported. Industry-clustered, heteroskedasticity robust standard errors are in parentheses below each estimated coefficient. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)
r_L	0.033 (0.028)	-0.001 (0.026)	0.029 (0.053)
$r_L \times trend$			0.0000 (0.0020)
r_S	0.667*** (0.078)	0.712*** (0.073)	0.315*** (0.100)
$r_S \times trend$			0.0110*** (0.0026)
r_A	0.720*** (0.080)	0.816*** (0.073)	1.380*** (0.145)
$r_A \times trend$			-0.0208*** (0.0056)
Fixed Effects			
# firms	3,061		3,061
# industries		66	
# quarters		175	
Observations	134,128	134,128	134,128

there can be differences across industries, and there could be differences in their accounting conventions. To be sure that we are truly measuring changes in asset price information and not these other confounding features, we can focus on the change in transaction efficiency associated with the implementation of Rule 19-b in May of 1975 and tighten the estimation window around this period.

I estimate panel regressions using the same framework as before, but I now interact the returns with a dummy variable, *post75*, that equals one for observations where all corresponding variables are observed after the advent of competitive commissions (i.e. after May of 1977). Interacting with this dummy variables tests for a discontinuity in the parameter estimates when crossing this boundary. This regression is reported in Table 3.6.

There are four regression specifications in the columns of the table, with each one representing a smaller window around 1975. The first specification estimates the panel regression over the full sample, comparing pre-1975 to post-1975 data using the observations from 1966 to 2010. Both long horizon and short horizon prices show dramatic increases in their information content, with their coefficients increasing by a factor of four. However, only the short horizon variables show statistical significance.

The three successive regression specifications with tighter and tighter sample windows increase the standard errors in the coefficient estimates but decrease the concern that other factors unrelated to efficiency and information are driving this result. Looking at the coefficient estimates, the post-1975 effect on short horizon price information remains roughly equal for each time window considered. The effect on long horizon information is always weaker than short horizon and difficult to

Table 3.6: Testing the May Day effect in the time series

The regression estimates below are the result of panel regressions of earnings news (Δx defined in section 3.3 of the paper) on past log returns and log returns interacted with a post-1975 dummy variable. Coefficients for constant term and constant post-1975 dummy are estimated but not reported. Industry-clustered, heteroskedasticity-robust standard errors are reported in parentheses. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	full-sample (1)	10 yr window (2)	5 yr window (3)	3 yr window (4)
r_L	0.010 (0.035)	0.010 (0.035)	0.016 (0.051)	0.069 (0.074)
$r_L \times post75$	0.031 (0.043)	0.078 (0.055)	-0.000 (0.062)	-0.137 (0.105)
r_S	0.234*** (0.066)	0.235*** (0.066)	0.255*** (0.086)	0.375** (0.153)
$r_S \times post75$	0.513*** (0.119)	0.560*** (0.180)	0.469** (0.208)	0.407 (0.319)
r_A	0.811*** (0.128)	0.812*** (0.129)	0.933*** (0.178)	0.870*** (0.251)
$r_A \times post75$	-0.124 (0.173)	0.514*** (0.191)	0.704** (0.287)	1.077** (0.496)
Fixed Effects # firms	3,058	1,653	1,205	1,059
Observations	128,114	55,184	30,160	18,070

distinguish from zero.

IDENTIFICATION USING CROSS-SECTIONAL COST DIFFERENTIALS

So far the panel analysis has only used the dimension of time to associate active trading and information with transaction efficiency. The strongest evidence for this channel will come from the differential impact across stocks.

The NYSE fixed commission schedule was always a function of the nominal share price. Assuming the nominal share price is a historical artifact, this creates variation across stocks that is plausibly unrelated to any economic characteristics. The commission schedule was set as a decreasing function of nominal share price, so stocks with lower prices were much more expensive to trade than those with higher share prices.¹¹

There are various ways to exploit this variation. The most simplistic is to use a difference in differences approach. I form three categories: *lowP* for stocks with a nominal share price less than \$15, *midP* for stocks whose nominal share price is between \$15 and \$30, and *highP* for stocks whose nominal share price is above \$30. We can then look at the differential impact across categories before and after 1975.

Table 3.7 reports the results of this approach, where the coefficients of interest are the magnitudes of the product: $r_L \times lowP \times post75$, $r_L \times midP \times post75$, $r_L \times highP \times post75$, $r_S \times highP \times post75$, and so forth. The prediction we are testing is whether these coefficients are positive (indicating more information post-1975) and

¹¹A surprising fact about stock prices is that the distribution of their nominal price per share has been remarkably consistent over time despite inflation and secular changes in investor and investment characteristics. This has been discussed by Weld, Michaely, Thaler and Benartzi (2009).

monotonically decreasing in nominal price (indicating a differential impact across firms according to the relative change in transaction efficiency). As in the previous table, each regression specification corresponds to tighter windows around 1975.

The results for short-horizon prices are just as predicted. All prices appear more informative, but the impact on securities with the largest change in transaction costs (*lowP*) is an order of magnitude higher than stocks where the change was more moderate. As hoped, the relationship is monotonic across the three categories and roughly consistent as the time window shrinks.

In the first regression specification, which uses the longest window, there is some evidence of an increase in information content of long-horizon prices, and the cross-sectional relationship with respect to nominal share price is monotonically decreasing. However, the statistical significance is low, and result disappears entirely in the specifications with shorter sampling windows.

3.4 IMPLICATIONS AND CONCLUSIONS

The empirical analysis shows great success in explaining the modern growth in the cost of capital markets and in looking at its effect on asset prices. However, looking at the information in asset prices only opens the door to broader questions about the social benefits of these changes.

In the simple model presented here, the benefits of active trading largely come from two sources: the noise shocks and the efficient allocation of capital. However, the improved capital allocation is a broadly shared positive externality, not something

Table 3.7: Testing May Day effect in the cross-section

Coefficients for constant term and unique permutations of constant dummies are not reported. Industry-clustered, heteroskedasticity-robust standard errors are reported in parentheses. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	10 yr window (1)	5 yr window (2)	3 yr window (3)
Long-horizon return			
$R_{LH} \dots$	-0.014 (0.020)	0.004 (0.029)	0.015 (0.042)
$\times lowP$	0.025 (0.120)	-0.009 (0.137)	-0.023 (0.180)
$\times midP$	-0.040 (0.038)	-0.087 (0.056)	-0.164 (0.111)
$\times lowP \times post75$	0.137 (0.191)	-0.091 (0.201)	-0.193 (0.261)
$\times midP \times post75$	0.125*** (0.047)	0.0639 (0.062)	0.113 (0.135)
$\times highP \times post75$	-0.013 (0.191)	-0.040 (0.287)	-0.0659 (0.496)
Short-horizon return			
$R_{SH} \dots$	0.186*** (0.040)	0.225*** (0.058)	0.313*** (0.089)
$\times lowP$	-0.0217 (0.225)	-0.027 (0.244)	-0.094 (0.325)
$\times midP$	0.131 (0.146)	-0.0375 (0.168)	-0.025 (0.209)
$\times lowP \times post75$	2.201*** (0.416)	2.056*** (0.520)	1.939*** (0.675)
$\times midP \times post75$	0.292 (0.241)	0.368 (0.248)	0.067 (0.292)
$\times highP \times post75$	0.164* (0.082)	0.144* (0.084)	0.170 (0.121)
Fixed effects			
# industries	64	61	61
Observations	61,198	36,174	24,084

the active investors accrue directly. The immediate trading profits come at the expense of a counterparty. To what extent will these noise traders be happy in funding trading profits?

SOCIAL WELFARE

The bigger normative question everyone wants to answer is: are we spending too much on finance? Taking the empirical results back to the modeling framework, we easily see two important welfare effects. First, investors fight over their slice of the pie, leading to what Stein (1987) terms "welfare-reducing speculation." These expenses are wasteful and would suggest too much spending in financial markets. Second, more informed asset prices increase the size of the pie, but the informed investors capture only a small portion of this benefit. All of us who use public market prices are free-riders, and this positive externality suggests we aren't spending nearly enough on informed speculation.

The welfare-reducing speculation can be clearly seen in the simple model where the supply of the risky investment is perfectly inelastic, as it would be for very short horizons. Using the same model parameters that illustrated the equilibrium in section 3.1, I add a dotted line to the left panel of Figure 3.8 to show the social welfare (calculated as average expected utility) in the same plot as the expected utility of the active and passive investors. Since the resources spent on information have no effect on total output, social welfare is maximized with practically no informed trading, a solution clearly less than the competitive equilibrium.

It is this type of intuition that drives the suggestions of Philippon (2010), who

suggests we may have too few engineers relative to financiers, or Bolton et al. (2011) who similarly contrasts an overabundance of financiers relative to entrepreneurs.

In contrast, the free-riding effect is illustrated in the case of an elastic investment supply, as we would expect for long horizons. The left panel of Figure 3.9 shows the equilibrium for the same parameters used in the previously discussed example, except the supply of investment will now respond to more accurate asset prices. As you can see, the socially optimal level of informed investment would allocate nearly half of investors to buy information, but the competitive equilibrium allocates far fewer since the uninformed investors are free riding on the social benefits of more informed asset prices.

This analysis builds on the fundamental insight of Hirshleifer (1971), who contrasts the private and social value of foreknowledge. In the model presented here, all information is foreknowledge, learning about information that will inevitably be public knowledge later.

CONCLUSIONS

In the aftermath of the recent financial crisis, scrutiny of financial institutions has increased. The growth in the resources poured into active investment and the surging compensation levels of financial professionals are used as *prima facie* evidence that financial markets have become inefficient, with many doubting that more active management leads to more informative asset prices.

In a stylized model, I show that investment research and trading are complements, which causes the quantity of both to increase. Financial markets become more

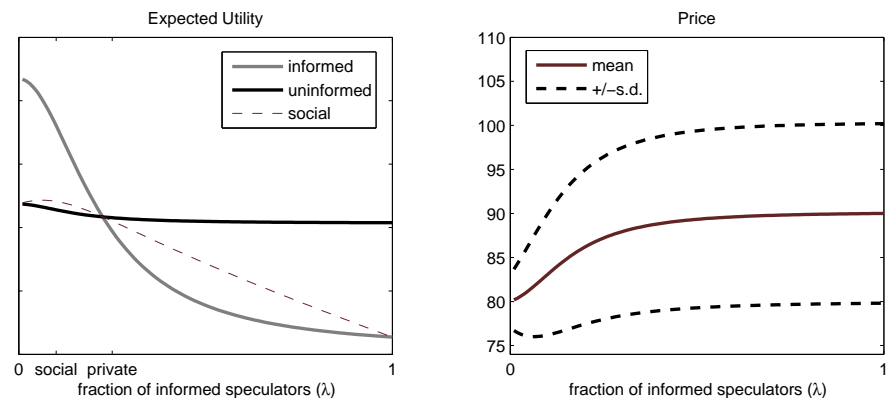


Figure 3.8: Welfare in the case of inelastic investment supply (short-horizon)

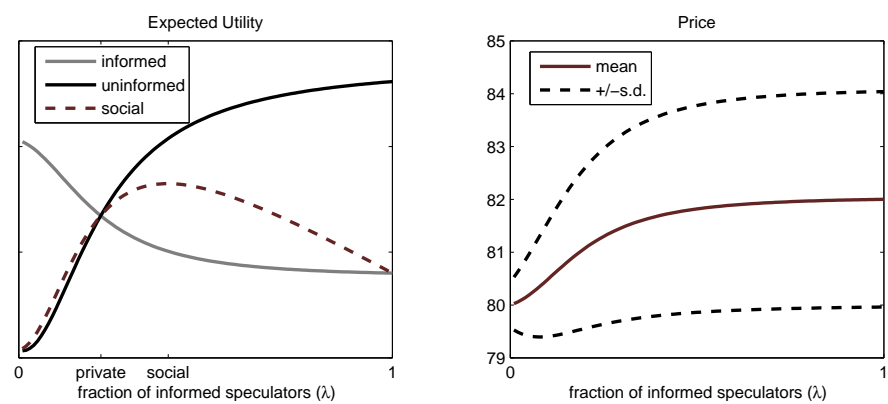


Figure 3.9: Welfare in the case of elastic investment supply (long-horizon)

informationally and operationally efficient. Empirically, this explanation is very successful in explaining the growth in resources spent in capital markets.

Furthermore, it introduces new evidence on the importance of time horizon. Trading horizons have shortened, and there is a corresponding increase in the short-horizon information contained in asset prices.

Since shorter trading horizons may not be socially optimal, this result could be interpreted as justification for Summers and Summers (1989) claim that a non-zero tax on trading might be welfare enhancing, although this requires more explicit measurement of the benefits that arise from informative markets and the recognition that the actual implementation of a financial transaction tax may be impractical (Campbell and Froot, 1994).

The types of dynamic strategies that become feasible with lower transaction costs not only make short-horizon information more valuable but they can also come closer to dynamically completing markets. It is certainly no accident that equity options became widely available in the late 1970s and early 1980s, precisely when US transaction costs experienced their largest drop. The newfound exposures made possible by dynamical hedging may have attracted investors to trade on new risks (Simsek, 2012).

The cost of capital markets has grown enormously over the past few decades. A portion of this can be attributed to the events of May 1975 that enabled dynamic trading strategies and spurred an increase in active investing. This opened the door to modern capital markets, with information and trades moving at ever shorter horizons.

References

- Adrian, Tobias and Joshua Rosenberg. December 2008. "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *Journal of Finance*, 63 (6), 2997–3030.
- Andersen, Torben G., Tim Bollerslev, Peter F. Christoffersen and Francis X. Diebold. "Volatility and Correlation Forecasting," in Graham Elliott, Clive W.J. Granger and Allan Timmerman, eds., *Handbook of Economic Forecasting, Vol I*, 2006. chapter Volatility, pp. 778–878.
- Antón, Miguel and Christopher Polk. 2010. "Connected Stocks," *Working Paper, London School of Economics*, (May).
- Bai, Jennie, Thomas Philippon and Alexi Savov. 2012. "Have Financial Markets Become More Informative?," *Working Paper, New York University*.
- Baillie, Richard T, Tim Bollerslev and Hans Ole Mikkelsen. September 1996. "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74 (1), 3–30.
- Bansal, Ravi and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59 (4), 1481–1509.
- Bansal, Ravi, Dana Kiku and Amir Yaron. "Volatility, the Macroeconomy and Asset Prices," 2012.
- Barberis, Nicholas and Andrei Shleifer. May 2003. "Style investing," *Journal of Financial Economics*, 68 (2), 161–199.
- Barberis, Nicholas, Andrei Shleifer and Jeffrey Wurgler. February 2005. "Comovement," *Journal of Financial Economics*, 75 (2), 283–317.

- Barndorff-Nielsen, Ole E and Shephard. June 2002. "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64 (2), 253–280.
- Beeler, Jason and John Y. Campbell. 2012. "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," *Critical Finance Review*, 1, 141–182.
- Berkowitz, Stephen A., Dennis E. Logue and Eugene A. Noser. 1988. "The Total Cost of Transactions on the NYSE," *Journal of Finance*, (1), 97–112.
- Black, Fischer. 1976. "Studies of Stock Price Volatility Changes," *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.
- Bollerslev, Tim. April 1986. "Generalized Autoregressive Conditional Heterskedasticity," *Journal of Econometrics*, 31 (3), 307–327.
- Bolton, Patrick, Tano Santos and Jose A Scheinkman. 2011. "Cream skimming in financial markets," *Working Paper, Columbia University*.
- Boyer, Brian H. February 2011. "Style-Related Comovement: Fundamentals or Labels?," *The Journal of Finance*, 66 (1), 307–332.
- Brainard, William C. and James Tobin. "Asset markets and the cost of capital," in "Economic Progress, Private Values and Public Policy: Essays in Honor of William Fellner," Amsterdam: North, 1977. pp. 235–62.
- Brealey, Richard A., Ian A. Cooper and Evi Kaplanis. December 2009. "Excess Comovement in International Equity Markets: Evidence from Cross-border Mergers," *Review of Financial Studies*, 23 (4), 1718–1740.
- Breeden, Douglas T. September 1979. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7 (3), 265–296.
- Calvet, Laurent E and Adlai J Fisher. October 2007. "Multifrequency news and stock returns," *Journal of Financial Economics*, 86 (1), 178–212.
- Campbell, John Y. June 1987. "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18 (2), 373–399.

- Campbell, John Y. 1993. "Intertemporal Asset Pricing without Consumption Data," *The American Economic Review*, 83 (3), 487 – 512.
- Campbell, John Y. 1996. "Understanding Risk and Return," *The Journal of Political Economy*, 104 (2), 298 – 345.
- Campbell, John Y. and Glen B. Taksler. December 2003. "Equity Volatility and Corporate Bond Yields," *Journal of Finance*, 58 (6), 2321–2350.
- Campbell, John Y and Kenneth A Froot. "International Experiences with Securities Transaction Taxes," in Jeffrey Frankel, ed., *The Internationalization of Equity Markets*, number January, Chicago, IL: University of Chicago Press, 1994. chapter 6, pp. 277–308.
- Campbell, John Y and Ludger Hentschel. June 1992. "No news is good news.," *Journal of Financial Economics*, 31 (3), 281–318.
- Campbell, John Y. and Robert J. Shiller. 1988. "Stock Prices, Earnings, and Expected Dividends," *The Journal of Finance*, 43 (3), 661 – 676.
- Campbell, John Y. and Tuomo Vuolteenaho. 2004. "Bad Beta, Good Beta," *The American Economic Review*, 94 (5), 1249 – 1275.
- Campbell, John Y, Christopher Polk and Tuomo Vuolteenaho. January 2010. "Growth or Glamour? Fundamentals and Systematic Risk in Stock Returns.," *Review of Financial Studies*, 23 (1), 305–344.
- Campbell, John Y, Martin Lettau, G Malkiel Burton and Yexiao Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *The Journal of Finance*, 56 (1), 1–43.
- Campbell, John Y., Sanford Grossman and Jiang Wang. 1993. "Trading Volume and Serial Correlation in Stock Returns," *Quarterly Journal of Economics*, 108 (4), 905–939.
- Campbell, John Y., Stefano Giglio and Christopher Polk. "Hard Times," 2012.
- Chacko, George and Luis M Viceira. 2005. "Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets.," *Review of Financial Studies*, 18 (4), 1369–1402.
- Chen, Joseph. "Intertemporal CAPM and the Cross Section of Stock Returns," 2003.

- Chen, Long and Xinlei Zhao. December 2009. "Return Decomposition.," *Review of Financial Studies*, 22 (12), 5213–5249.
- Christiansen, Charlotte, Maik Schmeling and Andreas Schrimpf. September 2012. "A comprehensive look at financial volatility prediction by economic variables.," *Journal of Applied Econometrics*, 27 (6), 956–977.
- Christie, Andrew A. December 1982. "The Stochastic Behavior of Common Stock Variances–Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 10 (4), 407–432.
- Cochrane, John H. June 1991. "Volatility tests and efficient markets A review essay," *Journal of Monetary Economics*, 27 (3), 463–485.
- Cochrane, John H. July 2008. "The Dog That Did Not Bark: A Defense of Return Predictability.," *Review of Financial Studies*, 21 (4), 1533–1575.
- Cochrane, John H. August 2011. "Presidential Address: Discount Rates.," *Journal of Finance*, 66 (4), 1047–1108.
- Cohen, Randolph B., Christopher Polk and Tuomo Vuolteenaho. 2009. "The Price Is (Almost) Right.," *Journal of Finance*, 64 (6), 2739–2782.
- Coval, Joshua D and Tyler Shumway. June 2001. "Expected Option Returns.," *Journal of Finance*, 56 (3), 983–1009.
- Dabora, Emil M. and Kenneth A. Froot. 1999. "How are stock prices affected by the location of trade?," *Journal of Financial Economics*, (2), 189–216.
- Daniel, Kent and Sheridan Titman. March 1997. "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns.," *Journal of Finance*, 52 (1), 1–33.
- Daniel, Kent and Sheridan Titman. 2012. "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review*, 1, 103–139.
- Davis, James L, Eugene F Fama and Kenneth R French. February 2000. "Characteristics, Covariances, and Average Returns: 1929 to 1997.," *Journal of Finance*, 55 (1), 389–406.
- De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers and Robert J. Waldmann. 1989. "The Size and Incidence of the Losses from Noise Trading," *The Journal of Finance*, 44 (3), 16.

- Duffee, Gregory R. August 2005. "Time Variation in the Covariance between Stock Returns and Consumption Growth," *Journal of Finance*, 60 (4), 1673–1712.
- Engle, Robert F. July 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50 (4), 987–1007.
- Engle, Robert F, Eric Ghysels and Bumjean Sohn. "Stock Market Volatility and Macroeconomic Fundamentals," 2009.
- Engsted, Tom, Thomas Q. Pedersen and Carsten Tanggaard. "Pitfalls in VAR Based Return Decompositions: A Clarification," 2010.
- Epstein, Larry G and Stanley E Zin. 1989. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57 (4), 937–969.
- Eraker, Bjørn and Ivan Shaliastovich. October 2008. "An Equilibrium Guide to Designing Affine Pricing Models," *Mathematical Finance*, 18 (4), 519–543.
- Eraker, Bjørn and Wenyu Wang. "Dynamic Present Values and the Intertemporal CAPM," 2011.
- Fama, Eugene F and Kenneth R French. 1988. "Permanent and Temporary Components of Stock Prices," *The Journal of Political Economy*, 96 (2), 246–273.
- Fama, Eugene F and Kenneth R French. February 1993. "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33 (1), 3–56.
- Fama, Eugene F and Kenneth R French. 1996. "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51 (1), 55–84.
- Francis, Jennifer, Katherine Schipper and Linda Vincent. August 2002. "Earnings Announcements and Competing Information," *Journal of Accounting and Economics*, 33 (3), 313–342.
- French, Kenneth R. 2008. "Presidential Address: The Cost of Active Investing," *The Journal of Finance*, 63 (4), 1537–1573.
- Glode, Vincent, Richard C. Green and Richard Lowery. 2012. "Financial Expertise as an Arms Race," *Journal of Finance*, (forthcoming).

- Greenwood, Robin. March 2008. "Excess Comovement of Stock Returns: Evidence from Cross-Sectional Variation in Nikkei 225 Weights," *Review of Financial Studies*, 21 (3), 1153–1186.
- Greenwood, Robin and David Scharfstein. 2012. "The Growth of Modern Finance," *Working Paper, Harvard Business School*.
- Greenwood, Robin and David Thesmar. July 2011. "Stock Price Fragility," *Journal of Financial Economics*, 102 (3), 490–471.
- Grossman, Sanford J. and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets," *The American Economic Review*, 70 (3), 393 – 408.
- Harvey, Campbell R. October 1989. "Time-Varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics*, 24 (2), 289–317.
- Harvey, Campbell R. March 1991. "The World Price of Covariance Risk," *Journal of Finance*, 46 (1), 111–157.
- Hirshleifer, Jack. 1971. "The Private and Social Value of Information and the Reward to Inventive Activity," *American Economic Review*, 61 (4), 561–574.
- Kandel, Shmuel and Robert F Stambaugh. February 1991. "Asset returns and intertemporal preferences," *Journal of Monetary Economics*, 27 (1), 39–71.
- Kumar, Alok and Charles M.C. Lee. October 2006. "Retail Investor Sentiment and Return Comovements," *The Journal of Finance*, 61 (5), 2451–2486.
- Labys, Torben G. Andersen. 2003. "Modeling and Forecasting Realized Volatility," *Econometrica*, 71 (2), 579 – 625.
- Lewellen, Jonathan, Stefan Nagel and Jay Shanken. May 2010. "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96 (2), 175–194.
- Liu, Jun. January 2007. "Portfolio Selection in Stochastic Environments," *Review of Financial Studies*, 20 (1), 1–39.
- Lo, Andrew W. and A. Craig MacKinlay. April 1990. "When are contrarian profits due to stock market overreaction?," *Review of Financial Studies*, 3 (2), 175–205.
- Merton, Robert C. September 1973. "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41 (5), 867–887.

- Murphy, Kevin M., Andrei Shleifer and Robert W Vishny. 1991. "The allocation of talent;," *Quarterly Journal of Economics*, (2), 503–530.
- Newey, Whitney K and Kenneth D West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55 (3), 703–708.
- Pástor, Luboš and Robert F Stambaugh. 2010. "On the Size of the Active Management Industry,," *Working Papers – Financial Institutions Center at The Wharton School*.
- Paye, Bradley. "Deja Vol: Predictive Regressions for Aggregate Stock Market Volatility Using Macroeconomics Variables," 2012.
- Philippon, Thomas. 2010. "Financiers versus Engineers: Should the Financial Sector be Taxed or Subsidized?," *American Economic Journal: Macroeconomics*, 2 (3), 25.
- Philippon, Thomas. 2012. "Has the US Finance Industry Become Less Efficient?," *Working Paper, New York University*, (May).
- Philippon, Thomas and Ariell Reshef. 2013. "Wages and Human Capital in the U . S . Financial Industry : 1909-2006," *Quarterly Journal of Economics*.
- Pindyck, Robert S. and Julio J. Rotemberg. 1993. "The Comovement of Stock Prices," *Quarterly Journal of Economics*, 108 (4), 1073–1104.
- Poon, Ser-Huang and Clive W J Granger. June 2003. "Forecasting Volatility in Financial Markets: A Review," *Journal of Economic Literature*, 41 (2), 478.
- Restoy, Fernando and Philippe Weil. January 2011. "Approximate Equilibrium Asset Prices," *Review of Finance*, 15 (1), 1–28.
- Roll, Richard. 1984. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39 (4), 1127–1139.
- Rosenberg, Tobias Adrian. 2008. "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *The Journal of Finance*, 63 (6), 2997 – 3030.
- Schwert, G William. December 1989. "Why Does Stock Market Volatility Change Over Time?," *Journal of Finance*, 44 (5), 1115–1153.

- Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends," *American Economic Review*, (3), 421–436.
- Simsek, Alp. 2012. "Speculation and Risk Sharing with New Financial Assets," *Working Paper, Harvard University*.
- Sohn, Bumjean. "Stock Market Volatility and Trading Strategy Based Factors," 2010.
- Stein, Jeremy C. 1987. "Informational Externalities and Welfare-Reducing Speculation," *Journal of Political Economy*, 95 (6), 1123–1145.
- Stein, Jeremy C. August 2009. "Presidential Address: Sophisticated Investors and Market Efficiency," *The Journal of Finance*, 64 (4), 1517–1548.
- Summers, Lawrence H. and Victoria P. Summers. 1989. "When financial markets work too well: A cautious case for a securities transactions tax," *Journal of Financial Services Research*, 3 (2-3), 261–286.
- Tauchen, George and Robert Hussey. 1991. "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59 (2), 371 – 396.
- Tinic, Seha M. and Richard R. West. 1980. "The securities industry under negotiated brokerage commissions:," *The Bell Journal of Economics*, 11 (1), 29–41.
- Veldkamp, Laura L. July 2006. "Information Markets and the Comovement of Asset Prices," *Review of Economic Studies*, 73 (3), 823–845.
- Weld, William C, Roni Michaely, Richard H Thaler and Shlomo Benartzi. 2009. "The Nominal Share Price Puzzle," *Journal of Economic Perspectives*, 23 (2), 121–142.
- Zhang, Andrew Ang. 2006. "The Cross-Section of Volatility and Expected Returns," *The Journal of Finance*, 61 (1), 259 – 299.